Synonymy and Lexical Priming
-- A Cross-Linguistic Investigation of Synonymy from Corpus and Psycholinguistic Perspectives

These submitted in accordance with the requirements

For the degree of Doctor of Philosophy

University of Liverpool

by

Juan Shao

29th March, 2018

The thesis is dedicated to

my loving and supportive parents

Acknowledgements

The most rewarding achievement in my life, as I approach middle age, is the completion of my doctoral thesis. This thesis took almost four years from conception to completion. It involved countless cycles of exploration, inquiry, meditation, enlightenment, doubt, confusion, uncertainty and perseverance. I feel I have learnt a lot from writing this thesis searching for the truth of science and life. This is the great treasure I will cherish not only in my future academic career but in my whole life.

This thesis would never have come to fruition without the support of many individuals. I would like to take this opportunity to express my immense gratitude to all those persons who have given their invaluable support and assistance.

I am profoundly indebted to my supervisor, Professor Michael Hoey, who throughout these many years has been very generous with his time and knowledge and assisted me at each stage to complete the thesis. He has given me a lifetime unforgettable memory of his intelligence, erudition, diligence, benevolence and patience.

I would also like to thank Dr. Hitomi Masuhara and Dr. Vittorio Tantucci for insightful professional advice. My research has benefited immeasurably from their experience, expertise and generosity.

Many colleagues of the English Department in the University of Liverpool have helped me as my co-researchers in the study, being always generous with their time to respond to me whenever I approached them for data and for discussion throughout the years of struggle and difficulty on this research.

Special thanks to Mr. Nicolas McGee and Ms. Stephanie Rose and all the staff and students from Wade Deacon High school for taking time to participate in this study without benefit to themselves.

My heartfelt gratitude especially to my parents, my brother and my sister-in-law. It was their love, understanding, support and encouragement that give me the strength and perseverance to overcome all difficulties and to finish this work.

Abstract

Synonymy and Lexical Priming: A Cross-Linguistic Investigation of Synonymy from Corpus and Psycholinguistic Perspectives

Juan Shao

With the development of computer technology and the availability of large corpora, recent linguistic studies have provided us with instances where looking at authentic language data has produced modifications of the way we think about language. A reconsideration of linguistic categories starts to emerge, with traditional terms being rejected or redefined. This research addresses the topic of synonymy from corpus and psycholinguistic perspectives both in English and Chinese to see whether we need to make modifications to the notion of synonymy.

The research starts with a psycholinguistic experiment to explore the psychological reality of synonymy. A word association test is carried out and the results show that people often do not have a shared sense of synonymy. On the one hand, people may offer various words as candidate synonyms for different types of prompt words. The words provided by the participants may be considered on occasion to be co-hyponymous, metonymous, or meronymous and or to be in a metaphorical relationship with the prompt words. On the other hand, there was found to be a relationship between candidate synonyms provided and the personal profile of the participants, including age, gender and subject field. The result of the psycholinguistic experiment seems to suggest that in people's minds the notion of synonymy exists but its boundaries with other semantic relations are sometimes unclear and synonymy is not a concept of clear-cut category.

To test whether a corpus approach can elicit similar findings to those of the psycholinguistic experiment, a corpus-driven analysis of eleven English candidate synonyms is carried out to test the validity of the notion of synonymy. It finds that the concept of synonymy is still usable but needs modification. Using a scale of similarity, we can only say that words are highly synonymous or synonymous to a certain degree. It is therefore concluded that well-established semantic relations such as synonymy, antonymy, hyponymy, metonymy and meronymy are helpful in talking about how words may be related to each other, but that it is not always possible, when looking at corpus data, to allocate a pair of words to one of these relations rather than another.

To test whether these findings for English are also true for Chinese, a case study comparing a pair of potential English synonyms with a pair of potential Chinese synonyms of equivalent meaning is first conducted to explore whether Chinese near-synonyms are primed differently in terms of their collocations, colligations, semantic associations and pragmatic associations. Then ten potentially synonymous words in Chinese are analysed. The results show that, as was found for English, the notion of synonymy is valid in discussions of Chinese lexis, but the boundary between synonymy and co-hyponymy is sometimes blurred. The similarities and differences between candidate synonyms both in English and Chinese could be identified with the categories utilised in lexical priming and the strength of synonymy among candidate synonyms could be measured by these categories.

Combining the findings of both the corpus analysis and the psycholinguistic experiment, the research shows that the notion of synonymy is more complex than we may think and that the ways people are primed may suggest possible explanations for the complexity of this linguistic phenomenon.

Table of Contents

Lists of Figures and Tables

Chapter 1 Introduction

1.1. Motivation for the study

There has been a long tradition of exploring the phenomenon of synonymy in linguistics, and a number of definitions and classifications of synonymy have been put forward, most of which focus on the degree of similarity and difference between lexemes in terms of their characteristic semantics or pragmatics. Making use of theoretical criteria, linguists such as Lyons (1968; 1977) and Cruse (1986) have provided classifications of synonyms. Lyons (1981) talks of three types of synonym: full, complete and total synonyms as a starting point for distinguishing absolute synonymy and partial synonymy. He proposes the notion of descriptive synonymy which he contrasts with complete synonymy. Cruse (1986) makes a further distinction between partial synonymy and near synonymy. He also introduces the notions of propositional synonymy and plesionymy. However, as Storjohann (2010) points out,

> Quite often, synonyms are considered to be words with sets of features that are formalized and attributed to logical relations, and most of these definitions were established at a time when linguists strove for a description of language as a system and attempted to show how vocabulary might be structured. (p. 70)

In semantics, synonymy is defined as a typical semantic relation in which two items have the same or similar meanings. Along with other semantic relations such as antonymy, co-hyponymy, metonymy and meronymy, synonymy has been neatly categorised. Although it has been recognised that synonymy is on a continuous scale (Cruse, 2002, p. 488) and that a strict categorisation of types of synonymy is problematic (Storjohann, 2010), the notion of synonymy itself does not seem to have been challenged.

However, there are two respects in which traditional approaches to synonymy might be challenged. The first is that the great majority of studies of synonymy look only at Indo-European languages, with examples taken overwhelmingly from English. The second is that the exploration and analysis of authentic language phenomena using corpora have challenged many other hitherto uncontroversial linguistic categories. This thesis will therefore look at potential synonyms not only in English but also Chinese, and it will make use of corpus linguistic methodology in its examination of such items.

With the development of computer technology, it has become possible to search for, retrieve, sort and make calculations from a large number of language data with high accuracy (Kennedy, 1998). The impact of corpora on linguistics has been compared to that of the telescope on astronomy. As Stubbs (1996) notes:

> The combination of computers, software, and large corpora has already allowed linguists to see phenomena and discover patterns which were not previously suspected. (p. 231-232).

Corpus research has called into question long-held beliefs about language, in particular the traditional breakdown into vocabulary and grammar. The research will start with a word association test designed to test whether people have an idea what synonymy is and then move on to a corpus approach to explore the validity of the notion of synonymy and also to test whether the findings in the psycholinguistic experiment and corpus analysis are consistent.

1.2. Does the concept of synonymy have psychological reality?

The concept of synonymy seem to have existed for a long time in linguistics; no empirical studies however appear to have tested the psychological reality of synonyms. The concept of similar meaning seems unproblematic; any native speaker of English comfortably recognises words such as *big* and *large*, *cold* and *freezing* as having similar meanings. There's no doubt that people have a receptive understanding of synonyms, but the question is not whether people can recognize synonyms, but rather whether they can produce synonyms on request. Therefore this thesis first seeks to test whether people have a shared sense of synonymy.

To explore the psychological reality of synonymy, a psycholinguistic experiment is first carried out. The purpose is to see how people understand the notion of synonymy. If the experiment supports the psychological reality of synonymy, it would be unnecessary to adopt the corpus approach to test the validity of the notion of synonymy. This test may serve as a preliminary stage for the later analysis to synonymy.

Now I will make an analogy between grammatical categories and synonymy to illustrate why both a multilingual perspective and a corpus approach may be worth adopting in exploring the validity of the notion of synonymy.

1.3. Why a bilingual approach to synonymy?

In traditional linguistics language is regarded as systematic, making use of clear-cut categorisations. Based on intuition and the introspection of linguists, grammatical categories (such as noun and adjective) and syntactic functions (such as subject and predicate) have been defined and illustrated with made-up examples. For example, noun refers to 'a word other than a pronoun that belongs to the word-class that inflects for plural, and that can function as subject or object in a sentence, can be preceded by articles and adjectives, and can be the object of a preposition' (The Oxford Dictionary of English

Grammar). In traditional school grammars, nouns have sometimes been defined notionally as a word that identifies a person, animal, place, thing, or idea. The notion seems easy to understand and the category seems to fit any language. No one would argue that *desk* and *car* are not nouns in English; 课桌 (*kè zhuō*) (desk) and 汽车 (*qì chē*) (car) can also be easily recognized as nouns in Chinese. In addition, a noun has certain features; for example in English, nouns can be categorised into countable and uncountable groups; countable nouns have both singular and plural forms. Nouns may also follow articles (*a* or *the),* or possessives (such as *my, his* or *their*) and function as the head of nominal groups serving as subjects or objects in sentences/clauses.

The introduction of other categories in traditional grammar follows this 'slot-and-filler' pattern. These categories seem to make sense and have been thoroughly described. However, Chinese challenges the distinctions between grammatical categories substantially, for example:

Example 1.1:学 习 是 件 苦 差事。

Xué xí shì jiàn kǔ chāi shì.

Study is PAR bitter thing.

Studying is a difficult thing.

As Chinese is non-inflectional and there is no article before the word, it is hard to say whether 学习 (*xué xí*) is a noun or verb. The distinction between grammatical categories seems to be neglected by many Chinese grammarians. For example, in *Modern Mandarin Chinese Grammar* (Ross and Ma, 2006), only common nouns (such as 水 *shuǐ*, 'water' and 思想 *sī xiǎng*, 'thought') and proper nouns (伦敦 *lún dūn*, 'London' and 长城 *cháng chéng*, 'The Great Wall') are introduced while other types of nouns are not even mentioned. In addition, dictionaries do not provide much useful information in distinguishing the part of speech of any particular word. In most Chinese dictionaries, the part of speech is not mentioned at all. Examples are The Xinhua Zidian (1st to 10th editions, 1982-2004) and the Modern Chinese Dictionary (1st-6th editions, 1978-2012). Only the Modern Chinese Standard Dictionary (1st edition, 2004 and 2nd edition, 2012) mentions the part of speech of words, stating for example that 学习 (*xué xí*) is a verb. In the book *A Practical Chinese Grammar for Foreigners*, Li and Cheng (2008) state that 'conversion of parts of speech' in Chinese is very common and further explain:

> 'If the meaning of a word in different sentences remains unchanged, it is not considered conversion although it has different functions. For example: 我们学习 [*We study*.] and 学习很重要 [*Studying is very important*.], in both sentences, 学习 is a verb' (p. 13).

However the explanation is very unconvincing and the English translation makes the situation even more confusing.

As a partly inflectional language, the identification of grammatical categories in English seems to be less controversial than Chinese. However look at the following examples:

Example 1.2. Her studying of Confucius inspired us to take up Chinese philosophy.

Example 1.3. The studying of Confucius is difficult.

Example 1.4. Lily studying Confucius inspired us to do the same.

Example 1.5. Studying Confucius is difficult.

Example 1.6. Studying is always difficult.

In example 1.2., the word *studying* shows some features of a noun: following the possessive *her*, being followed by a postmodifying prepositional phrase and functioning as the head of a group serving as subject in the sentence. However there are differences between *desk* and *studying*. Some people insist on labelling '*studying*' as noun because of the possessive *her*; others argue it is a nominalised verb; while still others call it a verbal noun or gerund (Houston, 1989). Example 1.3 is similar to example 1.2 in that *studying* can be considered as a noun following the definite article *the*. However compared with example 1.2, example 1.4 is missing *of* and *Lily studying Confucius* looks like a clause; consequently *studying* may be labelled as a verb despite the obvious similarities in use and meaning (and clausal positioning) to the instances in sentences 1.2 and 1.3. Again in example 1.5, s*tudying Confucius* looks like a clause and example 1.6 is the trickiest instance as *studying* could be either a verb or a noun.

To sum up, looking at Chinese shows that the distinction between nouns and verbs is not as straightforward as it appears to be in English. There is no easy way to distinguish nouns and verbs in Chinese and the distinction in English is not as neat as we sometimes assume. What this suggests is that we cannot take any concept for granted, nor the borders or boundaries of the concept. In our daily life, we are surrounded by unsystematic and sometimes even messy language phenomena. This thesis will consequently look at the viability of the concept of synonymy in both English and Chinese, and it will do so using corpus linguistic approaches.

1.4. Why a corpus approach to synonymy?

As we investigate phenomena using authentic data and modern corpus techniques, a reconsideration of linguistic categories starts to emerge. For example, on the basis of the analysis of authentic language data, Sinclair (1991a) demonstrates that there are few instances of '*of*' which are genuinely prepositions. Sinclair (1987) also talks of some of the main nouns and adjective classifications crumbling under corpus evidence and points out that 'even major parts of speech are not as solidly founded as they might be' (Sinclair, 1992). The COBUILD grammar (Sinclair, 1990) gives many examples where there is convergence of grammatical classes and lexical sets (Stubbs, 1996). In addition, building on his analysis of the way children are exposed to the numeral system, Hoey (2007) argues that the whole grammatical system is a product of collocational and other lexical patterns. Stubbs (2007) notes that 'empirical work on large corpora does not support a concept of fixed phrases, but rather of recurrent phrasal constructions, which are combinations of lexis and grammar, and which typically consist of a partly-fixed lexical core plus other variable items'. Therefore, he argues, language description should not be 'concept/definition to examples', but rather 'usage to concept'.

Stubbs' reflection on the concept of fixed phrases seems to suggest that the traditional language description may be challenged or at least modified with work on large authentic language data. The above discussion of grammatical categories both from a bilingual perspective and from a corpus perspective raises a series of questions about synonymy. First, if the boundary between nouns and other classes is difficult to define in Chinese and proves slightly less straightforward than sometimes assumed even in English, we may wonder whether the same is true of synonymy. Are there any situations where it is not possible to decide whether words are synonyms, just as it is sometimes difficult to say whether *studying* is a noun or verb? Secondly, since it can be shown that grammatical categories derived from English and other Indo-European languages appear to work less well in Chinese than in English, it may lead us to consider whether it is possible that the concept of synonymy also works less well in Chinese than in English. In other words, does the concept of synonymy work in the same way in different languages, especially those which are not part of the same language family? Finally, since a word association test has been conducted to explore the psychological reality of synonymy, could we find evidence from the corpus approach to support the findings in the psycholinguistic experiment?

Corpus-linguistic investigation of synonymy is a field little explored. Pre-corpus studies of synonymy focus on the logical description and identification of synonyms (Quine, 1953; Cruse, 1986; Lyons 1995; Edmunds, 1999) and early corpus approaches to synonymy only concentrate on the collocational and colligational behaviours of synonyms (Geeraerts, 1986; Divjak & Gries, 2006; Liu and Espino, 2012). Although some work has been done on differentiating synonyms and word choice in second language teaching, there have been relatively few cross-linguistic studies on synonymy, especially making use of languages with no family relationship, such as English and Chinese.

1.5. Aim and scope of the study

The research reported in this thesis starts with the aim of exploring the sense of synonymy in people's minds. First, the psychological reality of synonymy will be explored in an experiment. A word association test seems to be appropriate to elicit people's judgements on synonymy. If people are found to differ in their judgements, the causes of these differences will also be explored. Secondly, corpus analysis is conducted to see whether we could find evidence to support the findings in the psycholinguistic experiment. The analysis will start with a group of English words which are assumed to be synonymous and then corpus data will be used to test whether these words are really synonymous by looking at whether they share primings in collocation, semantic association and colligation when authentic language use is examined. Finally, if the corpus approach provides support for the notion of synonymy with English data, the study will continue to test the notion of synonymy by examining candidate synonyms in Chinese, again using a corpus approach.

The theory of Lexical Priming (Hoey, 2005) has been chosen as the framework of the study. Based on corpus analysis, Lexical Priming gives explanations for the existence of important phenomena unearthed by corpus linguistics including collocation, colligation and semantic association from a psycholinguistic perspective and the corpus-driven categories of descriptions which Lexical Priming utilizes are culture and language neutral because it has been shown that two typologically different languages (English and Chinese) share similar properties when looked at from both a lexical and psycholinguistic perspective (Hoey and Shao, 2015; see also Xiao & McEnery, 2006).

The thesis is concerned to answer the following research questions:

(1) How do people understand the notion of synonymy? Does synonymy have psychological reality? Do people share, or differ in, the meaning they ascribe to their sense of synonymy?

(2) If we find that synonymy has psychological reality, does the analysis of corpus data help to explain the findings obtained in my psychological experiment?

(3) Are the findings concerning synonyms derived from the analysis of Chinese data consistent with the findings concerning English synonyms derived from the same kind of analysis of English data? In other words, can we describe synonymy in the same way in both English and Chinese?

(4) If synonymy can be described in the same way in languages which have no family relationship, do the corpus-linguistic categories used by Lexical Priming enable us to identify similarities and differences between candidate synonyms in both English and Chinese?

(5) Given Cruse (2002)'s claims that synonymy is scalar, do the categories used in Lexical Priming help us to measure the strength of synonymy between pairs or among a group of words in the two unrelated languages?

(6) Based on the findings both from the psycholinguistic experiment and the corpus approach, can we justify the notion of synonymy, and if so how? Can we justify the continued use of the notion of synonymy and if so on what grounds?

To tackle these questions, as already noted, a word association test will be conducted and then the British National Corpus (BNC) and zhTenTen11 Chinese Corpus will be analysed with the Sketch Engine (Kilgariff, 2003).

1.6. Significance of the Study

One of the intended outcomes of the study is to explore the notion of synonymy from a psycholinguistic perspective (Chapter 4). Although both daily life experience and research study findings seem to support the psychological reality of antonyms, it does not follow that synonymy has a psychological reality amongst users of a language. This research will provide possible explanations of the complexity of the notion of synonymy that make use of both the theory of Lexical Priming and psycholinguistic data.

Another intended outcome of the study is at the theoretical level. Firstly, the research revisits the notion of synonymy and explores the relationship of synonymy to other semantic relations, for example, co-hyponyms, metonyms and metaphors. The traditional notion of synonymy will be modified and a corpus-driven approach to the notion of synonymy is proposed to characterise the features of synonyms (Chapter 5). Secondly, the study will make use of the categories in Lexical Priming to describe English synonymy. Hoey (2005) has provided evidence in support of the argument that 'synonyms differ in respect of the way they are primed for collocations, colligations, semantic associations and pragmatic associations' (p. 79). He draws on analyses of nouns only; this thesis seeks to expand the claim to apply to potentially synonymous English verbs, in other words, to see how lexical priming can help us identify similarities and differences between synonymous verbs in English (see Chapter 6). Thirdly, Lexical Priming is claimed not to be culture or language specific. Studies have demonstrated its application to other languages, for example German (Pace-Sigge, 2015) and Arabic (Salim, 2011); this study seeks to test its applicability to the Chinese language, that is, whether categories used in lexical priming enable us to describe the semantic behaviours of Chinese words (Chapters 6). Finally the study seeks to address synonymy across two languages which have no family relations. The cross-linguistic analysis of synonyms between closely related words in English

and Chinese in terms of collocation, semantic association and colligation may help our understanding of the similarities and differences between these two unrelated languages, and will explore whether we can use collocation, semantic association and colligation to describe synonymy in the same way in both English and Chinese (see Chapter 7).

The final intended outcome of the study is at the methodological level. Previous corpus studies on synonyms have usually started with a synonymous pair and looked at their similarities and differences (Divjak, 2006; Gries 2001; Cries & Otani, 2010; Liu & Espino, 2010). This study however starts with potentially synonymous items and conducts a corpus-driven analysis to see how a corpus approach can enable us to decide whether candidate words are synonymous or not. In addition, previous studies have focused on a pair of words or at most five or six potentially synonymous words; this study however analyses large groups of potentially synonymous items: namely eleven English candidate synonyms and ten Chinese candidate synonyms.

1.7. Structure of the thesis

This thesis is organised into four sections. The first section (comprising Chapters 2 and 3) sets up the scene for the current study by reviewing the development of corpus linguistic approaches to language description (Chapter 2) and the literature on synonymy (Chapter 3).

This section is followed by a psycholinguistic experiment explores the psychological reality of synonymy (Chapter 4) and provides possible explanations as to why people may offer different candidate words as synonyms, drawing on the theory of Lexical Priming as well as corpus data.

In the next section a corpus-driven analysis of potentially synonymous words in English is carried out with the purpose of revisiting the concept of synonymy and exploring corpus methods for characterising synonyms (Chapter 5).

Section 4 (Chapters 6 and 7) deals with synonymy across languages. The applicability of lexical priming to Chinese is tested with a pair of synonymous verbs (Chapter 6). The study is then extended in Chapter 7 with a corpus-driven analysis of a group of potentially synonymous words in Chinese and the results are compared with the findings for English in Chapter 5.

Chapter 8 summaries the conclusions of the thesis and suggests further avenues for research.

Chapter 2 Development of Corpus Linguistics

The next two chapters set the scene by considering corpus-based studies on synonymy from a cross-language perspective. Firstly, the history of corpus studies is briefly reviewed and key concepts in corpus linguistics are defined. Then following a review of research studies on synonymy both before and after the beginning of the corpus era, contrastive studies are introduced. Finally I will explore the possibility of studying English and Chinese synonymy within the framework of lexical priming.

2.1. Emergence and early stages of corpus linguistics

Corpus linguistics has undergone a remarkable development in the last forty years. From being a marginalized approach used in English linguistics, particularly in English grammar studies, corpus analysis is now 'increasingly multilingual' (McEnery and Wilson, 1996) and can be illuminating in 'virtually all branches of linguistics and language learning' (Leech, 1997).

Long before the discipline of corpus linguistics evolved into its current state, linguistics had a substantial corpus-based history. Corpus methodology may be dated back to the pre-Chomskyan period. Despite not identifying themselves as corpus linguists and using shoeboxes filled with paper slips and simple collections of written or transcribed texts, linguists in the early twentieth century used a basic methodology which would qualify as corpus-based because 'the approach … began with a large collection of recorded utterances from some language, a corpus. The corpus was subjected to a clear, stepwise, bottom-up strategy of analysis' (Harris, 1993, p. 27). A number of achievements were made in branches of language studies during that time. For example, Jespersen (1949) and Fries (1952) used paper-based corpora to study grammar.

As McEnery and Wilson (1996) have pointed out, the 'debate between rationalists and empiricists, triggered by Chomsky in linguistics, is a very old one. Any discipline may face the basic decision of whether to rely on naturally occurring observations or artificially induced observations'. Although the methodology was empirical and based on observed data, it was severely criticized by Chomsky largely because along with the practical problems of data processing and the potentially infinite nature of language, it was argued that a corpus (collection of texts) could never yield an adequate description of language. Chomsky was right when he made the criticism as at that time the size of 'shoebox corpora' was generally very small and 'skewed'.

However, Chomsky's claim that a corpus could never be a useful tool for linguists and that linguists or language experts could build an adequate language model based on his/her intuition or introspection is disputable. Intuition and introspection can be useful in language analysis but should be applied with caution. They can be influenced by one's dialect or sociolect. What appears to be right to one person

may be unacceptable to others. However, a corpus can test what is actually done and draw conclusions from that about grammaticality and acceptability with authentic or real texts. When checking whether native-speaker intuition is right in making a particular introspective judgment about language, corpora are useful resources of accurate language information. Complementing the findings that intuition on its own misses, a corpus can yield more reliable quantitative data. McEnery and Wilson (2007) have pointed out:

> Corpus-based observations are intrinsically more verifiable than introspectively based judgments. […] Not only does it seem that the corpus appears a rather more reliable source of frequency based data, it is also the case that it provides the basis of a much more systematic approach to the analysis of language (p. 14-15).

Therefore, corpus-based approaches are more advantageous than traditional intuition-based approaches, which rejected or ignored corpus data, in that they do not usually go to the extreme of rejecting intuition but rather find the balance between the use of corpus data and the use of one's intuition (McEnery et al, 2006).

However, not all linguists accept the corpus methodology. One of the criticisms is that corpora are 'skewed' as language is non-enumerable and therefore no finite corpus can adequately represent language. Chomsky (1962) states:

> Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are *impolite*. The corpus, if *natural*, will be so wildly skewed that the description [based on it] would be no more than a mere list (p. 159).

Another problem that early corpus linguists faced was one of data processing. Searching a corpus of million words by eye was very time consuming and error prone. Without data processing ability, corpus methodology was slow and expensive, inaccurate and therefore ultimately infeasible.

Fortunately with the development of computer technology it only takes a few minutes to carry out the process of searching for, retrieving, sorting and calculating linguistic data with high accuracy. Although great contributions have been made by manual analysis over centuries, the radical change in linguistics took place in the last third of the 20th century with the availability of digital computers (Kennedy, 1998, p. 5-6).

In addition, Chomsky's criticism on the skewedness of corpora helped to foster a more realistic attitude towards corpus building. Modern Corpus linguistics emerged in the 1960s when linguists were not satisfied with the ways languages were being described (Teubert, 2004). It took issue with the size, representativeness, balance, and sampling of the data described. Although some linguists still hold

reservations about or objections to corpora, corpus-based methods are these days used to study a wide variety of topics within linguistics (Biber et al., 1998). The following section briefly introduces the development of, and major projects in, modern corpus studies.

The first large-scale project to collect English language data kicked off in the late 1950s. Known as the Survey of English Usage (Quirk, 1959), it engaged in empirical research focusing on grammar rather than meaning. Even though the spoken component of the survey was the first to be computerized and transcribed and the spoken data made widely available, the project did not have much immediate impact on data-orientated research due to the pervasiveness of the Chomskyan paradigm.

Other data-orientated projects of importance included the Brown corpus compiled in the 1960s and the LOB (Lancaster-Oslo-Bergen)-corpus in the 1970s, composed of American texts and British corpus respectively. The two corpora were manually tagged with part-of-speech information. They did not attract much attention from American linguists perhaps because of the relatively small quantity of data in spite of the earlier expectation that questions concerning grammar and lexicon would be answered. The LOB-corpus was later exploited for grammar and word frequency, but not meaning.

Technical problems and issues with the standardization of corpus compilation led to a retreat from the mainstream in respect of a quest for meaning in corpus research from the 1970s till the 1990s. In the 1990s the size of corpora could reach tens of millions of running texts. The features of machine-readable electronic corpora and the development of language processing software packages have facilitated linguistic analysis and advanced our understanding of language. According to differences of purpose, representativeness, organization and format, different types of corpora were compiled such as general corpora, specialized corpora, training corpora, test corpora, dialect corpora, monitor corpora, synchronic corpora, and diachronic corpora (Kennedy, 1998, p. 19-20). As the compilation of large corpora and the consideration of the representativeness of corpus sampling have become more sophisticated, they have provided opportunities for more specialized work:

> With a high degree of accuracy of measurement, computers have facilitated quantitative studies in generalizations about language and language use, which have helped renew and strengthen links between linguistic description and various applications. (Kennedy, 1998)

Leech (1991) also comments that ‘neither the corpus linguist of the 1950s who rejected intuition, nor the general linguist of the 1960s, who rejected corpus data, was able to achieve the interaction of data coverage and the insight that characterise the many successful corpus analysis of recent years’.

In addition, qualitative analysis data are used for more than providing ‘real-life’ examples of particular phenomena. As Schmied (1993) has observed, a stage of qualitative research is often a precursor for quantitative analysis, and it is more useful to consider the approaches as complementary in corpus

linguistics. According to McEnery & Wilson (1996, p. 76), ‘qualitative forms of analysis offer a rich and detailed perspective on the data’ while ‘quantitative studies enable one to discover which phenomena are likely to be genuine reflections of the behaviour of a language and which are merely chance occurrence’ . Therefore, corpus linguistics benefits from combining quantitative and qualitative perspectives on the same research questions.

2.2. Key terms and main contributions in corpus linguistics

‘Now corpus linguistics is inextricably linked to the computer, which has introduced incredible speed, total accountability, accurate reliability, statistical reliability and the ability to handle huge amounts of data’ (Kennedy, 1998). The development of digital computers and software cannot alone improve the quality of corpus linguistic research; what also makes an impact is the theoretical approach that is adopted. European corpus linguistics has been much affected by the thinking of Firth (1951 et seq.) both with respect to data and terminology.. This section introduces some important concepts and major contributions related to my study that have come out of the Firthian tradition.

2.2.1. meaning and form

‘Traditions deriving from Bloomfield and early Chomsky have always had extreme difficulties in combining rigorous distributional analysis of language forms with a theory of meaning’ (Stubbs, 1996). For Chomsky (1957), ‘grammar is automatous and independent of meaning’.

Later work in generative semantics and work inspired by speech act theory took the debates in different directions, but did not solve the form-meaning problem (Stubbs, 1996).

Corpus linguistics is maturing methodologically and the range of languages addressed by corpus linguists is growing annually (McEnery and Wilson, 1996). However, the corpus linguistic studies which appeared in the 1990s did not devote much space to the study of meaning. In their short book Corpus Linguistics, Tony McEnery and Andrew Wilson (1996) did not consider the issue of meaning. Kennedy (1998) likewise only spent 10% of the content of his book *An Introduction to Corpus Linguistics* on ‘lexical description’ while Douglas Biber, Susan Conrad and Randi Reppen (1998) had just thirty pages on ‘lexicography’ in their book of similar size.

John Sinclair (1991) filled the gap, for example in his book *Corpus, Collocation, Concordance*. He detected that a word itself does not carry meaning, but that meaning is often made through several words in sequence. This is the idea that forms the backbone of much current corpus linguistics. Perhaps the most important early corpus project was English Lexical Studies and the first person who used a corpus specifically for lexical investigation was John Sinclair, the pioneer in taking up the novel concept of the

collocation, introduced by Harold Palmer and A. S. Hornby in their *Second Interim Project on English Collocations* (1933).

According to Sinclair:

> In all cases so far examined, each meaning can be associated with a distinct formal patterning…there is ultimately no distinction between form and meaning… [The] meaning affects the structure and this is … the principal observation of corpus linguistics in the last decade. (Sinclair, 1991a; 1991b)

Although working on a very small electronic text sample (by contemporary standards), Sinclair succeeded in modifying the traditional view of the word as the core unit, investigating the meaning of 'lexical item' and exploring the relationship between the word and the unit of meaning. According to Sinclair (2004), understanding a segment of text is not the result of accumulating the meanings of each successive meaningful unit, as the flow of meaning is not from item to text but from the text to the item, which he called 'semantic reversal'. Sinclair (2004, p. 135) points out: 'The effects of reversals can be seen in dictionaries and lexicons when a word is frequently found in collocation with another, and this has an effect on the meaning'. He then illustrated it with the example of 'white wine'.

> *White wine* is not white, but ranges from almost colourless to yellow, light orange or light green in colour. That is to say, the meaning of *white* when followed by *wine* is a different colour range from when it is not. Traditional dictionaries tend to obscure this point by using encyclopedic information to explain the meaning… [and it] assumes that the user already knows roughly what colour *white* is when collocated with *wine*. (p. 135)

Sinclair's approach to lexical meaning is inspiring and intriguing. In his later theory, he put forward that 'the form of a linguistic unit and its meaning are two perspectives on the same event' (p. 139).

For Sinclair (2004), 'the meaning of words together is different from their independent meanings'. He suggested that 'the word is not the best starting-point for a description of meaning, because meaning arises from words in particular combinations' (p. 148).

These days in corpus linguistics it is generally accepted that form and meaning are very closely related and that variation in one normally leads to variation in the other.

2.2.2. lexis and grammar

Traditional linguistics treats grammar and lexis as two separate systems, and 'has been massively biased in favour of the paradigmatic rather than the syntagmatic dimension' (Sinclair 2004, p. 140). Sinclair (2004) argues that 'this initial division of language patterning may not be fundamental to the nature of

language, but more a consequence of the inadequacy of the means of studying language in the pre-computer age' (p. 165). When coping with the large range of variation in language, traditional linguistics 'puts most of the variation to one side through the device of separating grammar and semantics at the outset. This then obscures most of the structural relevance of collocation and removes any chance of the precise alignment of form and meaning' (Sinclair, 2004, p. 140).

Corpus linguistics, with the aid of computer technology, endeavours to present the relation between form and meaning more accurately by keeping the balance between the two dimensions. The concept of lexico-grammar has long been proposed by Halliday (1961), but Sinclair's detailed lexico-syntactic studies take the argument further than Halliday's position that 'lexis is the most delicate syntax' (Stubbs, 1996). Sinclair (1992) provides a simple lexical example of co-selection of lexis and grammar in showing that the noun '*lap*' is more likely to occur in a prepositional phrase in adjunct position, than to occur in the subject or object of a clause.

In addition, Francis (1991) provides a more systematic demonstration of the phenomenon that all words have their own grammar. She takes a number of nouns from a specific frequency band of English (for example *context, darkness*) to check whether they would be evenly distributed over different grammatical positions in the clause: subject, object, indirect object, adjunct, qualifier, and so on. The result shows that the distribution of different lemmas in the same grammatical position is very uneven. For example, *context* and *darkness* are much more common in adjunct position than elsewhere, whereas *impact* and *independence* are much more common in object position (Stubbs, 1993).

The explicitly pedagogical Cobuild grammar (Sinclair, 1990), which associates structures with lexical items, is a stage towards a thoroughgoing lexico-grammar (Stubbs, 1996). Although the lists which provide lexical items with structures are incomplete, they already provide information which is not available from introspections.

In a paper on 'the nature of the evidence', Sinclair (1991a) discusses the lemma *SET* and in particular its uses in the phrasal verb *SET IN* to show that different forms of a lemma pattern differently (Stubbs, 1996). Of all the forms, *set* is more frequent than *sets* and *setting*. The past tense of its verbal uses *set* is commonest. *Set in* tends to occur in the end of clauses and its subjects usually have negative or unpleasant associations (Stubbs, 1996). In addition, by documenting the different patterning of 'eye' and 'eyes', Sinclair (1991b) shows the non-equivalence of singular and plural form of nouns. He states that 'there is hardly any common environment' between the two word forms and that they 'do not normally have the capacity to replace each other'. These works have provided analyses of various lemmas and given precise examples of co-selection of lexis and grammar (Stubbs, 1996).

Sinclair based his thesis on two main arguments: first, there is no essential difference between 'lexical words' (or 'content words') and 'grammatical words' (or 'empty words'); and secondly the observed patternings of lexical items are observations about lexis and grammar.

A model which reconciles the paradigmatic and syntagmatic dimensions of choice is set out by Sinclair (1991, 2004). Five categories of co-selection are put forward as components of a lexical item, including core and semantic prosody, which are obligational, plus collocation, colligation and semantic preference, which are optional. The next section will introduce the terms which are central to the current study.

2.2.3. collocation

The concept of collocation is one of the most essential in corpus linguistics. The British linguist J.R. Firth discussed it as early as 1951 and first coined the term in its modern linguistic sense along with the famous explanatory slogan: 'you shall judge a word by the company it keeps' (1957). According to Firth (1968), 'collocations of a given word are statements of the habitual or customary places of that word.'

Firth's notion of collocation is essentially quantitative (Krishnamurthy, 2000). Throughout his discussion of collocation, Firth (1957) states actual numbers of occurrences for words in Lear limericks as well as using expressions such as *habitual, commonest, frequently, not very common, general, usual* and *more restricted* (Krishnamurthy, 2000). Firth's statistical approach to collocation is accepted by many other corpus linguistics for example, Halliday (1996), Greenbaum (1974), Sinclair (1991), Hoey (1991), Stubbs (1995), Partington (1998), McEnery and Wilson (2001) and Hunston (2002).

Halliday (1966) identifies the need to measure the distance between two collocating items in a text. More importantly, he brings in the concept of probability, thereby raising the need for data, quantitative analyses, and the use of statistics. Greenbaum (1974) reserves the terms 'collocability' and 'collocable' for potential co-occurrence, using collocation and collocate solely for words which frequently co-occur. However, the definition does not tell us how frequent the co-occurrence of two lexical items should be to be considered as collocation. Hoey (1991) states 'the statistical definition of collocation is that it is the relationship a lexical item has with items that appear with greater than random probability in its (textual) context'. The random probability can be measured using statistical tests such as the MI (mutual information), t or z scores. Hunston (2002) argues that 'while collocation can be observed informally' using intuition, 'it is more reliable to measure it statistically, and for this a corpus is essential'. This is because a corpus can reveal such probabilistic semantic patterns across many speakers' intuitionS and usage, to which individual speakers have no access (Stubbs, 2001).

Writers on collocation have picked up different aspects of Firth's ideas. Sinclair, who was a student of Firth's at London University, sees collocation as 'the occurrence of two or more words within a short space of each other in a text' (Sinclair, 1991, p. 170) and describes collocation as 'an observable phenomenon in language made visible in concordances' (Pace-Sigge, 2013). Sinclair (1996) points out that the 'idiom principle' grows out of 'frozen collocations'. Stubbs (1996) describes collocation as 'syntagmatic relations between words as such, not between categories'. In addition, Pace-Sigge (2013) argues that 'collocations are more than words appearing together in one context. Once a statistically high frequency of use is established, they can be seen as more than just chunks of words but rather as meaningful clusters that have idiomaticity'.

It is Hoey who brings a psychological perspective to the discussion of collocation. He asks how collocation comes into being, and by quoting Leech (1974) and Partington (1998) he gives reasons why speakers would collocate:

> We can only account for collocation if we assume that every word is mentally primed for collocational use. As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context (Hoey, 2005, p. 8).

Psycholinguists highlight why there are collocations and not mere co-occurrence of words. They have constructed experiments over the past decades that prove that human minds connect some words more closely than others. For example, in Meyer and Schvanefeldt's (1971) experiment, candidates are presented with two strings of letters and asked to respond 'same' if both strings are words or non-words, otherwise responding 'different'. The response time for pairs of commonly associated words was shown to be decisively quicker than the one for unrelated terms, indicating that the reader/listener makes a subconscious mental connection between these two nodes (Pace-Sigge, 2013).

Halliday and Hasan also (1976) state:

> Without our being aware of it, each occurrence of a lexical item carries with its own textual history, a particular collocational environment that has been building up in the course of the creation of the text and that will provide the content within which the item will be incarnated on this particular occasion (p. 289).

However following Hoey, Pace-Sigge (2013) argues that

> It is not the creation of a text that makes us collocate. We carry, without being aware of it, a template in our heads to collocate certain words, and these subconsciously recognisable collocates create the sense of cohesion for the reader (p. 14).

To sum up, in addition to its statistically demonstrability and its observability in concordances, collocation also contributes to the 'naturalness of language' (Hoey, 2005) due to its psychological origins.

2.2.4. colligation

Colligation is put forward by Firth, who introduces it thus:

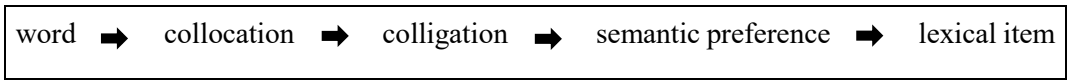
> The statement of meaning at the grammatical level is in terms of word and sentence classes or of similar categories and of the inter-relation of those categories in colligation. Grammatical relations should not be regarded as relations between words as such – between 'watched' and 'him' in 'I watched him' – but between a personal pronoun, first person singular nominative, the past tense of a transitive verb and the third person singular in the oblique or objective form. (Firth [1951]1957, p. 13)

However, the term was for a long time little used since being introduced. In the discussion of the lexical item 'naked eye', Sinclair (2004) observes that the pattern in L2 position (the second position to the left of *naked eye*) is dominated by two words (*with* and *to*) and other prepositions including *by, from, as, upon* and *than*, which account for over 90% of the concordance data, in which case Sinclair redefined the concept colligation as 'the co-occurrence of a grammatical class with a collocating pair' (in contrast to Firth's definition).

The relationship between collocation and colligation seems to vary. Based on the work of language use in context by Malinowski, Firth makes use of the term as follows:

> Colligation represents the syntactic juxtaposition of two or more grammatical categories. Colligation is derived from the concept of collocation which is the means of starting the 'meaning' of the word according to the habitual company it keeps; there is however no necessary relationship between colligation and collocation (Firth quoted in Bursill-Hall 1960, p. 247).

It seems that Firth regards colligation as standing independent of collocation (Pace-Sigge, 2013). The view however is not totally accepted by Sinclair and Hoey. Sinclair's discussion of lexical item *naked eye* seems to suggest a link between 'grammatical choice' and 'lexical necessity' (Pace-Sigge, 2013). Sinclair (1990) puts colligation in the middle of a continuum:



Inspired by Halliday (1959), Hoey (1997) divided colligation into two main classes: textual position and grammatical context. The former refers to a strong tendency that a lexical item may have to occur in a certain textual position more than others, e.g. at the beginning or end of a text. The latter refers to the way that a lexical item will tend to 'co-occur with a particular grammatical category of items'.

Building on Hoey, Susan Hunston (2001) gives a concise definition of colligation as 'the grammatical behaviour of a word in its various senses' and also states that 'there is no longer sense in distinguishing between lexis and grammar', therefore dissolving the relationship between collocation and colligation. In addition to affirming the viability of Hoey's ideas about colligation, Hunston (2001) points out that 'the phraseology of an individual text repeats the phraseology of innumerable other texts, and derives meaning from this repetition'. The evaluation foreshadows one of key ideas in lexical priming theory, namely that 'meaning lies in sequence of words and that meaning is created through repetition' (Pace-Sigge, 2013).

In addition, McEnery and Hardie (2012) state 'colligation is not simply a matter of co-occurrence with particular parts of speech; patterns of consistent co-occurrence of a word with different syntactic contexts are also described as colligation'.

Echoing Sinclair's approach to probability and frequencies, Stubbs (1996) highlights that 'strong probabilistic relations between lexis and syntax should find a place in grammar'.

In proposing lexical priming theory, Hoey (2005, p. 43) gives a tighter definition of colligation as follows:

> The grammatical company a word (or word sequence) keeps either within its own group or at a higher rank; the grammatical functions preferred or avoided by the group in which the word or word sequence participates; the place in a sequence that a word or word sequence prefers (or avoids).

It is important to note that Hoey extends colligational properties beyond a single word. According to Pace-Sigge (2013), Hoey's word sequence is close to Sinclair's lexical item and 'these sequences often (though not always) appear in the form of collocational clusters'.

Hoey (2005) also extends this by adding a concept of nesting, which 'implies a less linear, more cluster like relationship, in which collocations and colligations of the same sets of words can form different relationships' (Pace-Sigge, 2013).

2.2.5. semantic prosody, semantic preference and semantic association

The concept of semantic prosody was originally outlined by Louw (1993). It describes the characteristics of a word in terms of some aspects of its semantic context. The context has implications for the meaning of a word since the prosody becomes part of the word meaning (Starcke, 2008).

The term 'prosody' is borrowed from Firth (1957), who uses it to refer to phonological colouring which spreads beyond segmental boundaries. Rather than focusing on individual phonetic segments in terms of phonemes and allophones, Firth places a significant emphasis on how sounds work in a context to create meanings. He used the term 'prosody' for the many ways in which a sound may be influenced by its environment (McEnery and Hardie, 2012).

The notion of semantic prosody is intended to be directly parallel to this. Louw (1993) defines semantic prosody as '[a] consistent aura of meaning with which a form is imbued by its collocates' and argues that the habitual collocates of a form are 'capable of colouring it, so it can no longer be seen in isolation from its semantic prosody'. Prosodies are described by Louw (1993) as 'reflections of either pejorative or ameliorative [semantic] changes [over a period of time]' (p. 169) and 'based on frequent forms can bifurcate into good and bad' (p. 171).

Semantic prosody, also referred to as 'discourse prosody' by authors following Stubbs' (2001) usage, may be understood as a concept related to that of connotation in more traditional approaches to semantics. Partington (1998) refers to semantic prosody as 'the spreading of connotational colouring beyond single word boundaries' (p. 68). However, the key difference between semantic prosody and connotation is that the semantic prosodies are not necessarily accessible to intuition, which is often used to make judgments about the connotations of a word (McEnery and Hardie, 2012). Louw (1993) argues that a semantic prosody can only be discovered by analysis of a concordance.

A number of examples are given in the early literature of semantic prosodies. For example, Sinclair (1987, 1991) writes that items *happen* and *set in* are habitually associated with unpleasant events. Of *set in* he comments:

> The most striking feature of this phrasal verb is the nature of the subjects. In general they refer to unpleasant states of affairs. Only three refer to the weather; a few are neutral, such as *reaction* and *trend*. The main vocabulary is *rot* (3), *decay*, *ill-will*, *decadence*, *impoverishment*, *infection*, *prejudice*, *vicious (circle)*, *rigor mortis*, *numbness*, *bitterness*, *mannerism*, *anticlimax*, *anarchy*, *disillusion*, *disillusionment* and *slump*. Not one of these is desirable or attractive (Sinclair, 1987, p. 155-6).

Stubbs (1995) illustrates semantic prosody with the item *cause*, which, to an even greater extent than *happen*, carries bad news around with it, for example cancers are 'caused' much more frequently than cures. Other terms identified as having negative semantic prosodies include *utterly* (Louw, 1993), *undergo* (Stubbs, 2001), *occur, come about, take place* (Partington, 2004) and *persistent* (Hunston, 2007).

Partington (1998) points out

> Since [an] item is imbued with an 'unfavourable prosody', it cannot, in normal circumstance, be used in a favourable environment. A phrase like *good times set in* would be a highly marked, probably humorous use (p. 67).

However, in analysing 'naked eye', Sinclair (1999) argues that the expression has a semantic prosody of 'difficulty', which is clearly more semantically specific than mere negative evaluation, and he defines semantic prosody as 'attitudinal' and argues that semantic prosodies are more than merely positive or negative evaluation. The attitudinal focus that Sinclair incorporates into his use of the term perhaps triggers the parallel proposal of the separate concept of semantic preference. Stubbs (2001) defines semantic preference as 'the relation, not between individual words, but between a lemma or word-form and a set of semantically related words'. However, this definition seems to suggest a fuzzy boundary between semantic prosody and semantic preference. Xiao and McEnery (2006) show the closeness of the two terms semantic prosody and semantic preference.

Nevertheless one distinction between the two is that whereas a semantic preference may be a relation with any definable semantic field, a semantic prosody, in Louw's use of the term, is always a relation with either positive or negative evaluation (McEnery and Hardie, 2012). In addition, Sinclair (1999) observes that semantic prosody is 'on the pragmatic side of the semantics/pragmatics continuum. It is thus capable of a wide range of realisation'. McEnery and Hardie (2012) further explain that

> Semantic preference links the node to some word in its context drawn from a particular semantic field, whereas semantic prosody links the node to some expression of attitude or evaluation which may not be a single word, but may be given in the wider context (p. 138).

In spite of being the most widely used term in the literature (apart from collocation), semantic prosody and particularly Louw's account of it has been refined, questioned and criticised in various ways. Stubbs (2001) proposes the alternative term 'discourse prosody' and McEnery and Hardie (2012) suggest 'pragmatic prosody' on the basis that 'it is concerned with speaker meaning (pragmatics) rather than word meaning (semantics)'.

Hunston (2007) proposes the term ‘semantic preference’ or ‘attitudinal preference’ and points out that saying a word has a negative or positive semantic prosody involves taking a somewhat simplistic view of attitudinal meaning. She gives the example of *destruction*:

> The meaning is often not reducible to a simple positive or negative. It is essentially linked to point of view, so that there is often not one indisputable interpretation of attitude. (…) Destruction is a process which is often good for the destroyer but bad for the destroyed (p. 256).

Another example is the adjective *persistent*, which is a word that can be used to ‘indicate a mismatch of viewpoints, with the producer of a text indicating a difference between his or her own values and those of one of the participants in the text’ (Hunston, 2007, p. 256).

Based on Hunston’s findings, McEnery and Hardie (2012) point out that a degree of caution is required in making and evaluating claims that a particular word or phrase ‘possesses’ a particular semantic prosody.

Hoey (2005) also takes a different stand from Louw and Stubbs as the two terms ‘both seem to limit it to positive and negative effects’. He groups semantic prosody and semantic preference under the umbrella term ‘semantic association’ and admits that

> The terms semantic preference and semantic association may be seen as interchangeable. My reason for not using Sinclair’s term [semantic preference] is that one of the central features of priming is that it leads to a psychological preference on the part of the language user; to talk of both the user and the word having preferences would on occasion lead to confusion. (…) The change of term does not represent a difference of position between Sinclair and myself. (p. 24)

To summarise, all these key terms (collocation, colligation, semantic association etc.) will be made use of when we look at synonymy. The next section will review previous studies of synonymy both in English and Chinese, and also from a cross-linguistic perspective.

Chapter 3 Literature Review of Synonymy

3.1. Definitions and descriptions of synonymy

To define synonymy is never easy. The earliest attempt to define synonymy seems to start in ancient Greek philosophy, which laid a foundation for definitions and descriptions of synonyms. Philosophers focused on different aspects of synonyms, and the inconsistency in their definitions and descriptions reveal the controversy and complexity of synonymy.

The word 'synonymy' comes from ancient Greek 'syn' (with) and 'onama' (name), which might explain why definitions of synonymy involve expressions such as 'same names' or 'similar names'. However, the definition of synonymy is not as straightforward as we might expect. The earliest literature on synonymy seems to start in ancient Greek philosophy, which laid a foundation for definitions and descriptions of synonyms. Many fields including philosophy, psychology, applied linguistics, lexicography and more recently corpus linguistics and computational linguistic research have offered definitions, descriptions and analyses of synonyms, among which those in the field of philosophy are logical and analytic, and thus more abstract, while those in the other fields are more descriptive and empirical, often aiming to shed light on particular similarities, differences and usages.

Long before the existence of the term 'synonymy', some Greek sophists and philosophers in their dialogues and writings deliberately use synonyms in their texts to achieve effects like persuasion and cognition (Hüllen, 2004). The following section gives a brief review on how these philosophers focused on different aspects of synonyms. The inconsistency in their definitions and descriptions shows that the status of synonyms was controversial from the start and their description complex.

In one of his earliest dialogues *Laches* (Plato, 1953), Plato (428/7-349/7 BC) discusses the question of whether bravery should be a feature of the general education of boys. In the discussion Socrates and two military experts Laches and Nikias share the general presupposition that 'education must lead to virtues and that the virtues must serve the common good' (Borchert, 2006). The first suggestion raised in the discussion is that 'bravery is perseverance' and the second is that 'bravery includes a kind of knowledge which anticipates the future effects of one's own actions'. As the discussion goes on, the participants use a series of terms, which for linguists appear to be lexemes with overlapping and differentiating meanings. These lexemes include *(general) virtue, perseverance, bravery, courage, boldness, fearlessness, thoughtfulness, stupidity, justice, piety,* and etc. (Borchert, 2006). Because of their semantic affinity, some of the lexemes seem to be synonyms within a large semantic field.

Given the fact that synonyms prove to be practically important in the discussion that Plato reports, it is not surprising that other philosophers spend time trying to define synonyms. Democritus, a pre-Socratic philosopher, placed words into four categories: 'polyseme' (polúsēmon), different things are called by

the same name; ‘equality’ (isórrhopon, [today’s synonymy]), if different names will fit one and the same thing, they will also fit each other; ‘metonym’ (metōnumon), from the change of names; and ‘nameless’ (nōnumon), by the deficiency of similar items (Sluiter, 1993, p. 172-3).

However Democritus’ definition of isórrhopon (today’s synonymy) was countered by Prodikos of Keos, who showed that words commonly regarded as synonymous may in fact denote different things (Hüllen, 2004). According to Aristotle’s *Topics*, Prodikos compared synonymous lexemes systematically by explaining their semantic differences. For example, he juxtaposed positive and negative meanings (Mayer, 1913), by which he actually introduced the concept of antonymy as well. Moreover, he separated ‘essential (internal) features’ of word meaning from ‘accidental (external)’ ones, which foreshadows much later methods of dealing with synonymous words.

It is in the dialogue *Protagoras* (Plato, 1953) that Plato discusses the problems of synonyms more intensively. The topic is, as it happens, again the nature of (general) virtue and the problem of its teachability. Prodikos of Keo and another sophist Hippias of Elis (of whom we have no direct knowledge), are among the participants. The dialogue itself gives some examples of what Prodikos’ ideas about synonyms probably were.

Although the details are not available to us, it is obvious that language was already a central topic of philosophical discussions previous to Plato (Hennigfeld, 1994). This very selective and summary look at Prodikos and Plato shows that the awareness of semantic similarities between words goes back to the beginning of European thought. In various branches, Greek philosophy depended on the precise definition of terms. As terms cannot be expressed other than in other words, the linguistic method of determining word meanings with the help of related word meanings becomes the vehicle of concept discussion.

In Roman times, synonymy was dealt with indirectly in the vast programmes of cultural and linguistic education which were devoted to the arts of writing and oratory (Hüllen, 2004).

As a great thinker in the liberal arts tradition, Cicero inevitably dealt with synonymy in his writings. He was concerned with the proper language for the art of oratory, which had to ‘follow the postulates of correctness and of stylistic elegance’ (Borchert, 2006). He distinguishes between *loqui*, i.e. speaking in general, and *dicere*, i.e. the orator’s art of speaking, and emphasises that his way of writing depends on the particular situation. For example, in his letter to Lucius Papirius Paetus of October 46 BC, he writes:

> For I don’t always adopt the same style. What similarity is there between a letter, and a speech in court or at a public meeting? Why, even in law-case I am not in the habit of dealing with all of them in the same style. Private cases, and those petty ones too, I conduct in a more plain-spoken fashion, those involving a man’s civil status or his reputation, of course in a more

> ornate style; but my letters I generally compose in the language of every-day life. (*Letters to His Friends*, vol. II, bk. IX, sect. xxi; Cicero, 1965, p. 260-2, p. 261-3).

This reflection seems to point to an awareness of styles and registers, which prove important in later discussions of synonymy.

In the passages of *Quintilianus* (c.35-- c.100 CE), functions of synonyms in rhetorical ornament are discussed. According to Quintilianus, 'several words may often have the same meaning (they are called synonyms); some will be more distinguished, sublime, brilliant, attractive or euphonious than others' (p. 218-219). This opens a wide variety of usages for synonyms in various text genres (Hüllen, 2004).

The previous sub-section looked at how synonymy was dealt with by philosophers in the ancient Greek and Roman era. Their work laid the foundations for today's concept of synonymy. However, Hirsch (1975) points out 'the bulkiest literature on the subject of synonymy is to be found neither in literary theory, in linguistics, nor speech-act theory, but in analytic philosophy' (p. 562).

Since the late 1940s, a number of philosophers including Carnap, Quine, Lewis and Goodman have debated the possibility of synonymity (the philosophical term for synonymy). This sub-section attempts to summarise the main statements about synonymy in analytic philosophy and I will use the philosophical term 'synonymity' in the section.

Synonymity has been a major topic in philosophy since the publication of Rudolf Carnap's *Meaning and Necessity* in 1947, though it was discussed earlier in the writings W. V. Quine and C. I. Lewis.

Analytic statements in Quine's account fall into two classes:

(1) No unmarried man is married.

(2) No bachelor is married.

Quine (1953) regards the first statement as an acceptable notion of analytic truth. 'The relevant feature of this example is not merely true as it stands, but remains true under any and all reinterpretations of *man* and *married*. If we suppose a prior inventory of logical particles, composing *no, un-, not, if, then,* and etc. then in general a logical truth is a statement which is true under all reinterpretations of its components other than the logical particles' (Quine, 1953). The second statement is not a logical truth, for it does not remain true under every reinterpretation of its non-logical components 'bachelor' and 'married'. According to Quine, if (2) is nevertheless to be considered analytic, it is because we turn it into the logical truth (1) 'by replacing synonyms with synonyms' (Borchert, 2006). It seems that we can give an acceptable account of 'synonymity' in terms of interchangeability. However, this argument may raise the question whether a word and a phrase can be synonyms of each other. As we tend to use

different wordings each time we produce utterance, it may not be reasonable to consider a word and a phrase as being synonyms of each other.

One of the most widely discussed contributions to the topic of synonymity is Nelson Goodman's *On Likeness of Meaning*. Goodman (1952) proposes to explicate the notion of *synonymity* solely in terms of words and their 'extensions' – the object to which they apply. His account is confined to predicate expressions. He points out that 'we shall do better never to say that two predicates have the same meaning but rather that they have a greater or lesser degree, or one or another kind, of likeness of meaning… [And] their kind and degree of likeness of meaning is sufficient for the purposes of the immediate discourse' (p. 73).

In logical semantics (also referred to as analytical semantics), semanticists depend on synonymy in order to prove the truth of a statement. According to Miller & Charles (1991):

> Following a formulation usually attributed to *Leibniz* [referred to as the *salva veritate* principle], two words are said to be synonyms if one can be used in a statement in place of the other without changing the meaning of the statement (the conditions under which the statement would be true or false) (p. 1) .

Cruse (1986) states that 'the relation defined in terms of truth-conditional relations will be distinguished as propositional synonymy', which he defines and also provides an example for as follows:

> X is a propositional synonym of Y if (i) X and Y are syntactically identical, and (ii) any grammatical declarative sentence S containing X has equivalent truth conditions to another sentence S1, which is identical to S except that X is replaced by Y.
>
> An example of a pair of propositional synonyms is *fiddle* and *violin*: these are incapable of yielding sentences with different truth-conditions. For instance, *He plays the violin very well* entails and is entailed by *He plays the fiddle very well*. (p. 88)

In addition, dictionaries also provide definitions and descriptions of synonyms, for example:

> A synonym, in this dictionary, will always mean one of two or more words in the English language which have the same or very nearly the same essential meaning . . . Synonyms, therefore, are only such words as may be defined wholly, or almost wholly, in the same terms. Usually they are distinguished from one another by an added implication or connotation, or they may differ in their idiomatic use or in their application. (Webster's new dictionary of synonyms, 1984, p. 24)

> Strictly a word having the same sense as another (in the same language); but more usually any of two or more words (in the same language) having the same general sense, but possessing each of them meanings which are not shared by the other or others, or having different shades of meaning or implications appropriate to different contexts. (Compact Edition of the Oxford English Dictionary, 1989, V. II)

It is apparent that the definitions of synonymy as 'two or more words that mean the same' does not necessarily mean that 'they mean exactly the same'. Indeed the question arises whether 'true synonym does exist' and we also have to consider carefully what we mean by 'mean the same'. Goodman (1952) sparks a debate in his famous article *On Likeness of Meaning* by giving an example:

> [W]e cannot maintain the unqualified thesis that two predicates have the same meaning if they have the same extension [the set of things to which a concept or expression refers]. There are certain clear cases where two words that have the same extension do not have the same meaning. 'Centaur' and 'unicorn', for example, since neither applies to anything, have the same (null) extension; yet surely they differ in meaning. (p. 69)

Goodman (1952) further argues that:

> Although two words have the same extension, certain predicates composed by making identical additions to these two words may have different extensions. (p. 71)

Quine (1951) points out that

> Perfect synonymy -- whether understood as identity of meaning or identity of use -- is a logical impossibility… [T]o be able to say that two words 'have the same meaning' presupposes that we are able to contemplate meanings independently of the words used to represent those meanings. Since meanings do not come divorced from the meanings of their linguistic expression, to identify a synonym in terms of sameness of meaning is irredeemably circular. The only way out is to look for meaning in an expression's use. (cited in Taylor, 2003, p. 65)

Goodman (1952) finally concludes as follows:

> 1) No two different words have the same meaning;
>
> 2) There are no two predicates such that each can be replaced by the other in every sentence without changing the truth-value, even if we exclude all the so-called intensional contexts [e.g. All and only bachelors are bachelors/unmarried men];

> 3) [The definition of synonymy does not meet the requirement] that either of a pair of synonyms be replaceable by the other in all-non-intensional contexts without change of truth-value;
>
> 4) We shall do better never to say that two predicates have the same meaning but rather that they have a greater or lesser degree, or one or another kind, of likeness of meaning. (p. 69)

There are others who are for the proposition that synonyms do not exist. For example:

> It can…, be maintained that there are no real synonyms, that no two words have exactly the same meaning. Indeed it would seem unlikely that two words with exactly the same meaning would both survive in a language. (Palmer, 1981, p. 89)

The fact that terms such as near-synonym and approximate synonym have been coined is evidence that 'there is no such thing as a synonym' (Tognini-Bonelli, 2001).

3.2. Classifications and identification of synonymy

As discussed in the previous section, it is probably impossible to find two lexemes which have the exactly same meaning; hence the terms 'near-synonymy' or 'approximate synonymy' are advocated by many linguists. No matter which approach to synonymy is adopted, it seems to be accepted that synonymy refers to 'certain pairs or groups of lexical items [which] bear a special sort of semantic resemblance to one another' (Cruse, 1986, p. 265). As Cruse (1986) has pointed out, 'synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of implicit contrastiveness' (p. 266). This section discusses different approaches used to distinguish a word or phrase from its synonym.

According to Harris (1973), the traditional pastime of synonymists is to point out various ways of distinguishing between alleged synonyms. Collison (1939), for example, lists nine such ways:

> (1) One term is more general and inclusive in its applicability, another is more specific and exclusive, e.g. *refuse/reject*. Cf. *ending/inflexion, go on foot/march*.
>
> (2) One term is more intense than another, e.g. *repudiate/refuse*. Cf. *immense/great, towering/tall*.
>
> (3) One term is more highly charged with emotion than another, e.g. *repudiate* or *reject/decline*. Cf. *looming/emerging, louring/threatening*.

(4) One term may imply moral approbation or censure where another is neutral, e.g. *thrifty/economical, eavesdrop/listen*

(5) One term is more 'professional' than another; e.g. *calcium chloride/chloride of lime/bleaching powder; decease/death; domicile/house;* to *ordain* (*a priest*) or *induct* (*a vicar*), *consecrate or instal* (*a bishop*)*/appoint* (*a professor*).

(6) One term belongs more to the written language, it is more literary than another, e.g. *passing/death*. Within literary language further distinctions can be made, such as *poetical* and *archaic*.

(7) One term is more colloquial than another, e.g. *turn down/refuse*. The spoken language, too, includes further distinctions such as *familiar, slangy* and *vulgar*.

(8) One term is more local or dialectal than another, e.g. Scots *flesher/butcher*, or *to feu/to let*.

(9) One term belongs to child-talk, is used by children or in talking to children, e.g. *daddy, dad, papa/father* (in which different social levels are discernible), *teeny/tiny*, etc.

(Collison, 1939, p. 61-2)

Lyons (1981) posits three types of synonym: full, total, and complete synonyms, differentiating them on the basis of the totality of meaning and context. They are defined as follows:

(i) Synonyms are fully synonymous if, and if only, all their meanings are identical;

(ii) Synonyms are totally synonymous if, and only if, they are synonymous in all contexts;

(iii) Synonyms are completely synonymous if, and only if, they are identical on all (relevant) dimensions of meaning. (p. 50-1)

The three types are used as a starting point to distinguish 'absolute synonymy' and 'partial synonymy'. Lyons (1981) defines absolute synonymy as 'fully, totally and completely synonymous' and partial synonymy as 'synonymous, but not absolutely so' because they are either not complete 'on all (relevant dimensions of meaning)' or total. In other words they are not 'synonymous in all contexts' (p. 51). He also proposes the notion of 'descriptive synonymy', which he compares with 'complete synonymy' as follows:

> [T]he selection of one lexeme rather than another may have no effect on the message being transmitted. In this case, we can say that the intersubstitutable lexemes are completely synonymous. The selection of one rather than the other may change the social or expressive meaning of the utterance, but hold constant its descriptive meaning (if it has descriptive

> meaning) in which case, we can say that the intersubstitutable lexemes are descriptively synonymous (Lyons, 1977, p. 160).

In *Linguistic Semantics*, Lyons (1995) further distinguishes partial synonymy from near synonymy:

> Many of the expressions listed as synonymous in ordinary or specialized dictionaries (including *Roget's Thesaurus* and other dictionaries of synonyms and antonyms) are what may be called near synonyms: expressions that are more or less similar, but not identical, in meaning. Near synonymy, as we shall see, is not to be confused with various kinds of what I call partial synonymy, which meet the criterion of identity of meaning, but which, for various reasons, fail to meet the conditions of what is generally referred to as absolute synonymy. Typical examples of near synonyms in English are 'mist' and 'fog', 'stream' and 'brook', and 'dive' and 'plunge'. (p. 60-61)

In addition, Cruse (1986) presents various types of synonymy. Among those related to the current study are propositional synonymy (which was presented before), partial synonymy and plesionymy, each of which will be described below.

Cruse defines synonyms as:

> lexical items whose senses are identical in respect of 'central' semantic traits, but differ, if at all, only in respect of what we may provisionally describe as 'minor' or peripheral' traits (1986, p. 267).

He does not discuss what those 'minor' or 'peripheral' traits might be, but he presents the major traits as those dealing with register and collocation. He also illustrates partial synonyms in terms of grammatical collocation (what we refer as colligational differences in corpus linguistics) with the examples of *finish* and *complete*, where *finish* can be followed by a gerund, but not *complete*.

In talking about synonymy, Cruse (1986) distinguishes 'presupposed meaning' and 'evoked meaning'. According to Cruse, presupposed meaning is 'used in a pre-theoretical sense to refer to semantic traits which are taken for granted in the use of an expression, or lexical item, but not actually asserted, denied, questioned in the utterance in which they appear' (p. 278). Then he illustrates choice of synonyms due to selectional restrictions (i.e. 'semantic co-occurrence restrictions which are logically necessary') and collocational restrictions ('arbitrary co-occurrence restrictions'). He gives the following examples to illustrate:

> Example 3.1: Arthur died.

Example 3.2: ? The spoon died.

Example 3.3: Arthur kicked the bucket.

Example 3.4: ? The hamster kicked the bucket.

Example 3.5: ? The aspidistra kicked the bucket. (p. 279)

In examples 3.1 and 3.2, the verb *die* is semantically constrained by 'the nature of its grammatical subject', because only things that are organic, alive, and possibly mortal 'are logical prerequisites of the meaning of *die*' (p. 278). Therefore, it seems that Cruse's partial synonyms are constrained principally by what he terms selectional restrictions. With reference to examples 3.3 to 3.5, Cruse explains collocational restrictions as follows:

Unlike *die*, *kick the bucket* (in its idiomatic sense) is fully normal only with a human subject. But this additional restriction does not arise logically out of the meaning of *kick the bucket*. The propositional meaning of *kick the bucket* is not 'die in a characteristically human way', but simply 'die'; the restriction to human subjects is semantically arbitrary (p. 279). What Cruse fails to mention is that the collocational restriction of *kick the bucket* seems also related the origin of expression, as it started as a literal description of something that happens in the mechanics of hanging a felon and in this sense was of course never used of animals or machines.

Cruse (1986) points out that 'collocational restrictions vary in the degree to which they can be specified in terms of required semantic traits' (p. 280-281). He defines 'systematic collocational restrictions' as 'restrictions [which] behave as presuppositions of the selecting item' and refers to 'semi-systematic collocational restrictions' as the cases when the use of a particular lexical item 'sets up an expectation of a certain type of collocant, [though] there are exceptions to the general tendency' (p. 281) and gives the example of *client* and *customer* to illustrate as follows:

> A customer typically acquires something material in exchange for money; a client, on the other hand, typically receives a less tangible professional or technical service. Hence bakers, butchers, shoe-shops and newsagents have customers, while architects, solicitors and advertising agencies have clients. (p. 281)

Lastly, 'idiosyncratic collocations' concern items whose collocations 'can only be described by listing permissible collocants'. Cruse gives *flawless* as an example and shows that *flawless* could collocate with *performance*, *argument* and *complexion*, but not with *behaviour*, *kitchen*, *record*, *reputation*, or *credentials*.

In addition to presupposed meaning, Cruse (1986) introduces evoked meaning, another basis on which partial synonymy can be defined and classified. He explains that 'the possibility of evoked meaning is a consequence of the existence of different dialects and registers within a language'. Therefore dialectal synonyms can be created as the result of geographical (e.g. *autumn* and *fall*), temporal (e.g. *settee* and *sofa*) and social variations (e.g. *scullery*, *kitchen* and *kitchenette*). Register is another difference that would account for the choice of one synonym over another, which Cruse (1986) distinguishes in terms of three dimensions of variation: field, mode and style. Field refers to:

> the topic or field of discourse: there are lexical (and grammatical) characteristics of, for instance, legal discourse, scientific discourse, advertising language, sales talk, political speeches, football commentaries, cooking receipts, and so on. (p. 283)

Mode is concerned with 'the manner of transmission of a linguistic message – whether, for instance, it is written, spoken, telegraphed, or whatever' (for example, *concerning* is only used in written language and *about* in speech) and style refers to 'language characteristics which mark different relations between the participants in a linguistic exchange' (p. 284). Cruse points out that

> Style is of particular interest because this dimension of variation spawns the most spectacular proliferation of cognitive synonyms. The multiplication of synonyms [pertaining to style] is most marked in the case of words referring to areas of experience which have a high emotive significance such as (in [English] culture), death, sex, excretory functions, money, religion, power relations, and so on. For referents in these areas we typically find a range of subtly differentiated terms, which allows an utterance to be finely tuned to its context (p. 284).

After discussing propositional synonymy and partial synonymy, Cruse (1986) introduces 'plesionyms' and explains that:

> Plesionyms are distinguished from cognitive synonyms [propositional synonyms] by the fact that they yield sentences with different truth-conditions: two sentences which differ only in respect of plesionyms in parallel syntactic positions are not mutually entailing, although if the lexical items are in a hyponymous relation there may well be unilateral entailment. There is always one member of a plesionymous pair which it is possible to assert, without paradox, while simultaneously denying the other member. (p. 285)

Following are some examples that Cruse uses to illustrate plesionyms, which have occasionally been confused with near-synonyms.

> Example 3.6: It wasn't *foggy* last Friday − just *misty*.

Example 3.7: He is by no means *fearless*, but he's extremely *brave*.

Example 3.8: She isn't *pretty*, but in her way she is quite *handsome*.

Example 3.9: He was not *murdered*, he was legally *executed*. (p. 285)

According to Cruse, plesionyms cannot mutually entail; in other words, although there seems to be some overlapping in meaning, they cannot be substituted for each other.

Adding to the list, Edmunds (1999) divides synonyms (which he refers to as 'variation') into four categories: stylistic, expressive, denotational and collocational. The following table shows the variation types (with examples) for each category.

Classification of lexical variation with examples

Variation category	Variation type	Example
Stylistic	Geographical dialect	*loch : lake*
	Temporal dialect	*lapidate : stone (to death)*
	Social dialect	*loo : toilet*
	Language	*verboten : forbidden*
	Sublanguage	*matrimony : wedlock : marriage*
	Formality	*pissed : drunk : inebriated*
	Force	*ruin : wreck : destroy :*
	Concreteness	*name : christen*
	Floridity	*house : habitation*
	Euphemism	*toilet :bathroom : washroom*
	Familiarity	*divulge : disclose : reveal : tell*
	Simplicity	*hound : dog*
Expressive	Emotive	*daddy : dad : father*
	Expressed attitude	*skinny : thin : slim*
Denotational	Denotation	*account : chronicle : report*
	Implication	*mistake : slip : lapse*
	Suggestion	*help : aid : assist*
	Frequency of expression	*version : story*
	Fine-grained technical	*alligator: crocodile*
	Abstract dimension	*seep : drip*
	Continuous dimension	*mistake : error : blunder*
	Binary dimension	*escort : accompany*
	Complex 'dimension'	*begin : start : initiate*
	Specificity	*eat : consume : devour : dine : gobble*
	Extensional overlap	*high : tall*
	Fuzzy overlap	*forest : woods*
Collocational	Selectional restrictions	*land : perch*
	Idiom	*bite the dust : ~gnaw the powder*
	Grammatical collocation	*correct : right*
	Subcategorization	*teach : instruct*

(adapted from Bawcom, 2010, p. 25)

To summarise, no matter what terms linguists may use to refer to lexical items with the same or similar meaning, interchangeability/substitution seems to be one of the persistent criteria in identifying

potential synonyms. In addition to the linguists already cited, linguists who have adopted this criterion include Firth, 1951; Bolinger, 1975; Leech, 1974; Palmer, 1981; Lyons, 1981; Cruse, 1986; Hoey, 1991; 2005; Sinclair, 1991; and Stubbs, 2001. In dictionaries and thesauri, a number of synonyms may be offered circularly as the explanation to the entry word; however, these words may not always be used/interchanged in the same contexts.

3.3. Issues concerning definitions and descriptions of synonymy

In the modern study of synonymy, the focus is not on the link between language and reality, but rather on pieces of language which denote same/similar meanings.

> Two items are synonymous if they have the same sense. (Lyons, 1968, p. 428)

> Synonymy is used to mean 'sameness of meaning' (Palmer, 1976, p. 88)

From the definitions above, it can be seen that as with discussions of antonymy which usually refer to a pair of words having opposite meaning, we traditionally also define synonymy as existing between two items. It however needs reconsidering how many lexical items we should consider in defining synonyms, in other words, whether we have to confine synonymy to being a pair of words sharing similar meaning, or whether a list of words can be considered as synonymous. In fact, when asked to offer an antonym for a lexical item, people tend to provide one item, which indicate antonyms are usually grouped in pairs (Jones, 2006). Unlike with antonymy, typically more than two synonyms may be elicited from informants or given in dictionaries.

Whether 'items' or 'predicates' (Goodman's term) are words, phrases or sentences are another issue we need consider in defining or describing synonymy. Semantics traditionally recognises two main divisions: lexical semantics and phrasal semantics (Cruse, 1986). Lexical semantics studies word meaning, whereas phrasal semantics studies the meaning of phrase and sentence. For the current purpose, I will distinguish lexical meaning, phrasal meaning and sentential meaning, and their links to synonymy with examples of each following.

Example 3.10: He lives in a *big/large* house.

Example 3.11: *Due to/because of*

Example 3.12: Let's meet tomorrow morning.

I'll see you tomorrow morning.

Shall we meet tomorrow morning?

Why don't we meet tomorrow morning?

In example 3.10, *big* and *large* are the kind of synonyms I will be focusing on, which I refer to as 'lexical synonyms (synonymy between individual lexemes)' (Riemer, 2010). The phrases *due to* and *because of* in example 3.11 are referred as 'phrasal synonyms (synonymy between expressions consisting of more than one? lexemes)' (Riemer, 2010).

Talking about the two utterances '*I just felt a sharp pain*.' and '*Ouch!*' (Cruse, 1986, p. 271), Cruse (1986) claims 'there is a sense in which the content of the message conveyed by these two utterances is the same, or at least very similar' but they differ in what he calls 'semantic mode', the first being in 'propositional mode' and the second in 'expressive mode'. In addition, Partington (1998) gives the examples (a) *You make me sick.* and (b) *Will you ever grow up?*, and explains that 'the two utterances could perform the same function -- the same speech act, to use Austin's (1962) terminology -- presumably that of insulting or putting someone down'. Partington (1998) points out that 'this kind of synonymy is highly context dependent' and calls it 'illocutionary synonymy'. Sentences in 3.12 are the instances of illocutionary synonymy.

This thesis will however concentrate on lexical synonymy, which has been variously defined in the semantics literature. For some authors synonymy is a 'context-bound phenomenon' whereas for others it is 'context-free' (Riemer, 2010). According to Riemer (2010),

> Speakers do not characteristically seem to base their judgements of synonymy on a 'bottom-up' analysis of meaning of each of the words involved, concluding the words are synonymous if their separately established meanings are identical. Instead, a top-down procedure often seems to be at work: the fact that two expressions have the same contextual effect is what justifies labelling the substituted words as synonyms in that context. (p. 151)

In fact, many authors have considered substitution to be a criterion for synonymy. For example, Divjak et al. (in press) state that 'two words are considered synonymous in a sentence or linguistic context if the substitution of one for the other does not alter the truth value of the sentence. Two lexical units would be absolute synonyms if and only if all their contextual relations were identical'. However, it has been pointed out that no two items could be substituted in all contexts. For this reason, it is commonly asserted that absolute, perfect or full synonyms do not exist. No matter how close the meanings of two lexemes are, there are no absolute synonyms in reality. Therefore synonyms refer to lexical items where their senses 'are identical in respect of central semantic traits, but differ in respect of minor or peripheral traits' (Divjak et al., in press).

Antonyms, or words with opposite meanings, seem to be also very common in our daily life and speakers of English can readily agree that words like *good-bad, love-hate* and *in-out* are opposites or antonyms. Jones (2002) points out that 'recognising antonyms seems to be a natural stage in an infant's linguistic development' and he argues that 'our exposure to antonyms is not restricted to childhood; we are surrounded by 'opposites' throughout our adult life and encounter them on a daily basis' (Jones, 2002). In spite of the fact that antonyms are common it is not easy to identify the types and features of antonymy. Based on his analysis of newspaper corpus data, Jones (2002) identifies new classes of antonyms and demonstrates various features of them. Being usually grouped with antonymy, the types and features of synonymy however remain unknown. Based on my analysis of a small amount of language data, I provisionally sub-categorise lexical synonyms into four types: denotational synonyms, conceptual synonyms, contextual synonyms and metaphorical synonyms. Examples of each type are the following:

Example 3.13: *mist/fog*

Example 3.14: *idea/concept, purpose/aim*

Example 3.15: I'll tell my *big/elder* sister.

I live in a big/*elder house.

Example 3.16: *fruit/result* of research

The first type of lexical meaning illustrated in 3.13 usually involves concrete objects/actual beings which we can see, touch or feel in real life. Whether *mist* and *fog* are normally considered as synonyms is not the focus here, but rather they are given as an example of a candidate pair of denotational synonyms. As to the examples in 3.14, we cannot see or touch them, but we have a concept in our mind that we can use in a comparison; thus they can be considered as 'conceptual synonyms'. In 3.15, *big* and *elder* could be considered as synonyms in the context of 'sister', but may not function as synonyms in other contexts, so the term 'contextual synonyms' is proposed for them. In 3.16, *fruit* can be used metaphorically in a way that might be regarded as synonymous to *result*; to these I would assign the term 'metaphorical synonyms'.

3.4. Synonymy and other semantic relations

From the above discussion it is clear that synonymy is a type of semantic relation between lexical items since it involves at least two lexical items. This section will discuss other semantic relationships and compare them with synonymy.

3.4.1. hyponymy and synonymy

Hyponymy is the lexical relation expressed in English by the phrase 'kind/type/sort of' (Reimer, 2010). A chain of hyponyms describes a hierarchy of elements, for example in 3.17 *pigeon* is a hyponym of *bird* since pigeon is a type of bird, and *bird* is a hyponym of *animal* as bird is a type of animal.

Example 3.17: animal: bird: pigeon, crow, eagle …

Example 3.18: Occupation: architect, policeman, teacher, tutor, trainer…

Under the semantic label of *bird, pigeon, crow* and *eagle* are called co-hyponyms. When the meanings of co-hyponyms are close, we can have pairs that function as synonyms, for instance, in 3.18 under the semantic label *occupation*, we have *architect*, *policeman*, *teacher*, *tutor* and *trainer*, among which the meanings of *teacher, tutor and trainer* are so close that they can be labelled as synonyms or 'similonyms' (Bawcom, 2010).

Taylor (2003) talks about the example of *eat* and its synonyms as follows:

> Generally speaking 'to eat' means 'to put food into one's mouth', whereas in the following phrases it means this in a particular specification each of which can be expressed by a synonym: to *eat* ice cream (to *lick*), to *eat* soup (to *swallow spoonful*), to *eat* a steak (to *chew*), etc.

Note here the examples Taylor gives may not be considered as synonyms by some people; the explanation seems to also suggest that the words *lick*, *swallow* and *chew* can be considered as a type of *eating something*, thus hyponyms of *eat* and co-hyponyms to each other. Therefore Taylor's example of *eat* and its synonyms seems to suggest the boundary between synonymy and co-hyponymy could be blurred sometimes. Another example is *story* and *fiction* (also see Chapter 4), as some people think they are synonyms while others consider them hyponyms.

3.4.2. metonymy, meronymy and synonymy

Metonymy refers to a semantic relation in which a thing or concept is called not by its own name but rather by the name of something associated in meaning with that thing or concept (Wikipedia: https://en.wikipedia.org/wiki/Metonymy). Meronymy refers to the semantic relation when we use a

word for the part to replace the whole, or the whole for the part, for example *hand* and *arm*, *seed* and *fruit*, *blade* and *knife*; and conversely *arm* is a holonym of *hand* (Riemer, 2010). Look at the following examples.

Example 3.19: use your head

Example 3.20: lose one's head

The word *head* in 3.19 refers to the *brain*, a part of the head, arguably the organ which we use for thinking or the ability to think. In either case *brain* may be considered as a meronym of *head*. However, it is also possible to argue that *head* is used as synonymous to *brain*. In 3.20, *head* refers to *mind* or *ability to reason*, in which case it would be possible that *head* is treated as a metonym of *mind*, or *head* and *mind* are treated as synonyms. Therefore it can be argued that sometime the boundary between meronymy/metonymy and synonymy is not clear-cut.

3.4.3. metaphor and synonymy

According to Lakoff (1993), a metaphor refers to 'a novel or poetic linguistic expression where one or more words for a concept are used outside its normal conventional meaning to express a similar concept'. On the traditional view, metaphor is seen as a matter of literary use which asserts a resemblance between two entities (Riemer, 2010). Lakoff and Johnson (1980) observed that all the expressions in example 3.21 can be labelled as '*obligations are physical burdens*'. Though the underlying idea in each expression is different, they 'all essentially make reference to the same similarity between obligation and physical burden' (Riemer, 2010).

Example 3.21:

a. She's loaded with responsibilities.

b. She shouldered the task with ease.

c. She's weighed down with obligations.

d. She's carrying a heavy load at work.

e. I have to get out from under my obligations.

f. I have a pressing obligation.

g. She bears the responsibility for the success of this mission.

h. We shouldn't overload her.

(Reimer, 2010, p. 247)

Lakoff and Johnson (1980) claim that 'the very idea of obligation is conceptualized through the idea of a physical burden' and refer to it as 'conceptual theory of metaphor'. Riemer (2010) points out that the theory 'focuses on metaphors as a cognitive device which acts as a model to express the nature of otherwise hard-to-conceptualize ideas' (p. 247).

Example 3.22:

The first *fruit* of their work was legislation which provided that no land which was not already operational could become so unless certain planning requirements were met.

On the conceptual metaphor view, in 3.22 the concept of *result*, *outcome* or even *achievement* is set up with correspondence to the easily understood thing, *fruit* of plant. Therefore the word *fruit* in 3.22 is used metaphorically as *result* or *achievement*.

However, as metaphors enter into our everyday speech and lose their allusiveness and novelty, they become 'fossilised' or 'dead'. Some may argue that *fruit* in 3.22 has lost its metaphoricity and become 'literal' in daily use. The focus here however is not on whether the metaphor is 'dead' or 'living', but rather on whether it is possible to consider the words *fruit*, *result* and *achievement* to be synonyms in this context.

3.4.4. antonymy and synonymy

Antonymy and synonymy are more common semantic relations than hyponymy, metonymy and meronymy and usually considered to be easily distinguishable from each other. However, in some cases the boundary between the two may also be blurred. Partington (1998, p. 31) mentions that 'close synonyms are frequently treated as opposites, or at least as being in some sort of opposition'. He gives the example *No, it's not roasted, it's boiled*, in which *roasted* and *boiled* are put in a relation of oppositeness.

3.4.5. polysemy and synonymy

The term 'polysemy' is used for a word – or, to be more precise, a lexeme – that has two or more related senses (Tsiamita, 2011). The relationship between polysemy and synonymy is different from other

above-mentioned semantic relations. The above section has discussed the fuzzy distinctions between synonymy and other semantic relations, but there is no way we confuse polysemy and synonymy as polysemy refers to one lexeme having multiple meanings while synonymy is a semantic relation involving at least two items. The discussion of polysemy here is to address a methodological issue in studying synonymy from a corpus approach.

Most words are potentially polysemous. The fact that many linguists pointed out that relatedness of meaning is a matter of degree raises the question of how related two (or more) senses need to be to still be considered as belonging to a single lexeme. Different dictionaries may list different number of senses for the same word or lexeme. Gibbs & Matlock (2001) raise the possibility that 'lexical networks might not necessarily be the best way to describe polysemy' (p. 234), namely that

> all meanings of polysemous words might be tied to very specific conceptual knowledge and lexico-grammatical constructions as opposed to being encoded in a network form in a speaker's mental lexicon. This idea is consistent with the idea that there may not be strict, or even any, boundaries between the grammar and the lexicon. (p. 235)

The theory of Lexical Priming suggests just that: a blurring of the boundaries between the grammar and the lexicon to the point of a reversal 'of the roles of lexis and grammar, arguing that lexis is complexly and systematically structured and that grammar is an outcome of this lexical structure' (Hoey, 2005, p. 1).

The issue needs to be considered is of the distinction between synonymy of words and synonymy of senses. Remier (2010) gives the example of *pupil* and *student* and explains that

> *pupil* is arguably synonymous with *student* with respect to one of its senses (person being instructed by a teacher); but with respect to the sense 'centre of the eye' the two words are, of course, non-synonymous. (p. 152)

Murphy (2003) demonstrates that the pair *baggage/luggage* are synonymous with respect to the sense 'bags' but not with respect to the metaphorical sense 'emotional encumbrances'.

> Example 3.23:
>
> Check your baggage/luggage with the gate agent.
>
> I won't date guys with baggage/*luggage from their divorces.

In these cases we are dealing with polysemy, the case of a word having two or more meanings/senses. According to Hoey (2005), 'the collocations, semantic associations and colligations a word is primed for will systematically differentiate its polysemous senses' (p. 81). Hoey's (2005) observation has led to his 'drinking problem' hypotheses:

1. Where it can be shown that a common sense of a polysemous word is primed to favour certain collocations, semantic associations and/or colligations, the rarer sense of that word will be primed to avoid those collocations, semantic associations and colligations. The more common use of the word will make use of the collocations, semantic associations and colligations of the rarer word but, proportionally, less frequently.

2. Where two senses of a word are approximately as common as each other, they will both avoid each other's collocations, semantic associations and/or colligations.

3. Where either (1) or (2) do not apply, the effect will be humour, ambiguity (momentary or permanent), or a new meaning combining the two senses. (p. 82)

As a couple of studies have been conducted on testing these hypotheses with different language data (Hoey, 2005; Pace-Sigge, 2015), the focus here in not on providing more evidence. What seems to be related to the current study of synonymy is that if one word has several senses, their collocations, semantic associations and colligations with different senses may influence the statistical significance of attempts to identify the synonyms.

Cambridge dictionary online provides two senses for *consequence* as follows:

› a result of a particular action or situation, often one that is bad or not convenient;

› *of little/no consequence*, also *not of any/much consequence*: not important

Therefore, when we look at synonymous English words via a corpus-driven approach, it is possible that polysemous senses of the words (such as *consequence* and *fruit*) may compromise attempts to measure the strength of similarities among the candidate words.

3.5. Approaches to identifying synonymy

From the above discussion, it can be seen that scope and range have never been defined clearly in the definition of synonymy and there is no agreed terminology. Therefore issues arise when we attempt to identify synonyms, particularly in the situation that the boundary between synonymy and some other semantic relations is not very clear. Substitution/interchangeability and componential analysis are the most commonly used approaches in identifying synonyms, but they pose some problems. In the next

section, I will review the two approaches to argue they may not be reliable in identifying synonymy all the time.

3.5.1. synonymy and substitution/replaceability/interchangeability

Among the approaches used in the recognition of synonyms, substitution seems to be one of the most persistent criteria (Palmer, 1981; Lyons, 1981; Cruse, 1986; Sinclair, 1991; Stubbs, 2001). Dictionaries and thesauri often offer a number of synonyms circularly as the definition for each other; however, as discussed above, these words may not always be substitutable for each other in different contexts. More examples follow:

Example 3.24: *a big city/ a large city;*

a big pan/ a large pan

Example 3.25: *a big surprise/ a *large surprise;*

*a big success/ a *large success*

In example 3.24, *big* and *large* could be substituted/replaced, so it is safe to say *big* and *large* are synonyms. However, in example 3.25, we could say *a big surprise* and *a big success* but not *a *large surprise* or *a *large success*, therefore the two words *big* and *large* are not substitutable/replaceable, which suggests synonyms are more contextual than fixed.

Another criterion that has been suggested for identifying equivalence in meaning between words is that of signalling constructions such as *is, known as, called, that is, i.e.* and *or*. (Pearson, 1998). We therefore now discuss the functions of these signal words or phrases in identifying synonyms. For example:

Example 3.26: *To be afraid is to be scared* (Pearson's example)

In 3.26, *is* indicates a certain equivalence in meaning, so *afraid* and *scared* can be considered as synonyms. However, consider the following:

Example 3.27: *A tiger is a big cat.*

Example 3.28: *To see is to believe.*

Apparently, we cannot conclude that *tiger* is synonymous to *cat* or *big cat* in example 3.27; in this case *cat* is used as a generic term and hence is the superordinate of *tiger*. Some may argue that this is due to the unparalleled forms on either side of *is*, with *tiger* on the left being a word and *big cat* being a phrase. However if we look at example 3.28, we find that, even though they are used in perfect parallel in the structure, *see* and *believe* are definitely not synonyms according to anybody's definition. So this structure does not fulfil the task of unambiguously identifying synonyms.

Turning now to the other signal words/constructions frequently-used to identify synonyms, Pearson (1998) looked at connective phrases including *i.e., e.g., called, known as, the term* and *(*)* in the ITU corpus, GCSE corpus and Nature corpus. The analysis of the three corpora revealed that 'when certain phrases were present, it was sometimes possible to conclude that the words or phrases which co-occurred with these were in some way equivalent, whereby equivalence includes relations of synonymy, paraphrasing and substitution.' However, 'in many situations where the connectives phrases are apparently being used to denote a relation of equivalence, they are in fact functioning as connective phrases of genus-species relations'. Examples are:

Example 3.29: *The ability to simulate motion (**i.e.** animation) is a potential enhancement that can be achieved by several mean*s (from ITU corpus)

Example 3.30: *cell types, **e.g.** root-hair cell, egg cell (ovum), sperm cell, muscle cell, skin cell, leaf cell* (from GCSE corpus)

Example 3.31: *alternatively a single piece of equipment **called** a transmultiplexer can be used to perform the functions* (from ITU corpus)

Example 3.32: *A function which provides the user with the means to control system functions via MML inputs and outputs; also **known as** an IT function* (from ITU corpus)

Example 3.33: *surface uplift (**The term** is used to mean that the average elevation of the ground increases) on a regional scale is difficult to demonstrate* (from Nature corpus)

(All the examples here are from Pearson, 1998)

Although these signalling constructions can be used to identify equivalence in meaning to some extent, it is not always reliable.

3.5.2. synonymy and componential analysis

Componential analysis was developed in the second half of the 1950s and the beginning of the 1960s as an efficient way of analysing meaning. Kempson (1977) defines it as 'the meanings of words analysed not as unitary concepts but as complexes made up of components of meaning which are themselves semantic primitives' (p. 18). Componential analysis, also known as lexical decomposition, involves the analysis of the sense of a lexeme into its component parts (Lyons, 1995). Violi (2001) explains as follows:

The meaning of each term can be analysed by a set of meaning component or properties of a more general order, some of which will be common to various terms in the lexicon. There may [sic] in the lexicon. There may also be specific restrictions, for instance the nature and structure of features, and the procedures by which they are selected. However, the term componential analysis is often used to refer not only to simple decomposition into semantic components, but to models with much more powerful theoretical assumptions. (p. 53)

In structural semantics words are considered to be configurations of a number of meaningful components, which are called 'semantic features' and are given semi-formalised names, for example:

man: + *HUMAN* + *ADULT* + *MALE*

woman: + *HUMAN* + *ADULT* – *MALE*

HUMAN, ADULT and MALE are the 'semantic features' we could use to distinguish the compositional meaning of words *man* and *woman*. The symbols (+ & –) are used to indicate whether the word has this semantic feature or not.

Leech (1974) and Kempson (1977) both draw heavily on componential analysis in their analyses of antonymy and this strategy is effective when dealing with certain antonymous pairs, especially those which concern kinship terms or gender. Jones (2002) however has pointed out that 'the explanatory power of componential analysis does not seem to extend beyond this – describing an antonymous pair such as *bachelor/spinster* is unproblematic, but tackling a pair such as *active/passive* creates many more difficulties'(p. 12).

By demonstrating the same components, componential analysis may help us understand synonymy; for example, *adult* and *grown-up* share the semantic features of [+HUMAN] and [+ADULT]. In establishing degrees of synonymy, componential analysis can identify the similarities and differences

by indicating whether the word has certain semantic feature or not. For example, *barn* and *shed* have some but not all semantic components in common.

barn: + *BUIDLING* BUILDING + *STORAGE* +*FRAM* FARM +*FOR CEREALS* –*HOUSE*

shed: + *BUILDING* + *STORAGE* – *FARM* –*FOR CEREALS* + *HOUSE*

Pustejovsky (1996) however points out

> The act of defining 'componentiality' presupposes the act of decomposing. It is an analytical process. Indeed, the practice of defining content words usually takes the form of a decomposing enumeration of their parts (features). (p. 39–60)

It is difficult to decide which categories of semantic feature should be included, especially for those which are not actual objects but refer to conceptual existence, for example, *belief* and *faith*, which do not have component parts that can be enumerated. To sum up, both substitution and componential analysis have been useful in differentiating different semantic relations to some extent, but they are inadequate as criteria for synonymy. The next section therefore considers the corpus approach to synonymy.

3.6. Previous studies on near-synonyms in English

As discussed above, synonymy is hard to define and different classifications of synonymy may be adopted. The discrimination of near synonyms has always been a very challenging issue for linguists, lexicographers, dictionary-makers and language teachers in both L1 and L2 teaching. Philosophers, linguists and language teachers have approached synonymy from various perspectives. As noted in the previous section, most research into synonymy in the fields of philosophy and semantics has been analytic and mainly based on linguists' intuition or introspection, with a particular focus on describing and classifying synonyms. This section will review studies on synonyms that adopt a different perspective both before and after the development of corpus linguistics.

Harris (1973) looks at the links between synonymy and the linguistic analysis of natural language and explores what any native speaker thinks s/he is claiming when s/he claims that one expression is 'exactly synonymous' with another. However, his focus is on 'the theoretical consequences of supposing that a correct linguistic analysis of a natural language may, in certain cases, treat as identical in meaning two sentences – or, more generally, two items of whatever grammatical status – not identical in form' (p.

1). Although writing before corpus linguistics had established itself, he refers to the need for 'distributional criteria' and for the 'quantification' of synonymy. .

Adopting a cognitive perspective, Hüllen (2009) discusses the reasons why synonymy is an essential concept in lexical semantics. He states:

> 'Synonymy is a basic phenomenon of lexis because words can only be semanticized by words, which means that every word in a language has its synonyms. Besides, the rules of textual constitution demand that there be perfect synonyms to avoid repetition. … On the level of the system, so-called synonyms are still different from each other. But in performance and within given bounds, which are delimited by the lexemes, the meanings of words adopt certain senses following the constraints of co-texts and contexts. … In performed language -- not in the system created out of reflection -- words can therefore also adopt perfect synonymy.' (p. 145)

He explained these ideas in detail with illustrative examples. However, he also points out that '[the] deliberations are not corpus-based; rather they provide the guidelines for later work with corpora which I recommend strongly'.

It is indeed such empirical corpus studies that bring a new perspective to synonymy. At this point, we could conduct studies of synonymy with a corpus linguistics approach of the kind described in the previous chapter.

Most of the early corpus approaches to synonymy focused on the collocational and colligational behaviours of near-synonyms. For example, Geeraerts (1986) and Justeson & Katz (1995) found that the most effective way to disambiguate synonymous adjectives was to examine their noun collocates, that is, the nouns the synonymous adjectives typically modify. In addition, a number of corpus-based behavioural profile (BP) studies have been conducted on synonymous verbs (Divjak, 2006; Divjak & Gries, 2006; Hanks, 1996) and synonymous adjectives (Gries, 2001; Gries & Otani, 2010; Liu 2010). Using the corpus of Contemporary American English (COCA), Liu and Espino (2012) conduct a behavioural profile analysis of four near-synonymous adverbs *actually*, *genuinely*, *really* and *truly*. Their analysis shows that all four adverbs emphasize reality/truth and hence the central force pulls the adverbs together and makes them synonymous, but they differ from one another in varying degrees in their semantic functions. Based on this, Liu and Espino (2012) point out that due to the unique nature of adverbs the key usage features for the analysis and understanding of these lexical items are not all the same as those for the analysis and understanding of adjectives and verbs.

In addition to these studies of collocational behaviour, differences in the semantic prosodies of near synonyms are also explored, e.g., *fickle* is shown to be negative whereas *flexible* is shown to be positive (Tognini-Bonelli, 2001). Wen (2007) compares the semantic prosody of two near synonyms, *rather* and

fairly, based on analysis of the LOB (Lancaster-Oslo/Bergen) corpus and finds that though *rather* and *fairly* have the same denotational meaning, their semantic prosody differs from each other distinctively, as, in the adv + adj/adv colligation, *rather* tends to collocate with negative words like s*uperfluous, dismal, squalid, ugly, sad, sordid,* and *disappointing*, while *fairly* tends to collocate with positive words like *typical, safe, rapid, accurate, clearly, good, wide* and so on.

Some studies have been conducted on factors involved in the choice of synonyms. Wang and Hirst (2010) point out that in the context of near-synonymy, the process of lexical choice becomes profoundly more complicated. This is partly because of the subtle nuances among near-synonyms, which can arguably differ along an infinite number of dimensions. 'Each dimension of variation carries differences in style, connotation, or even truth conditions into the discourse in question' (Cruse, 1986), all making the seemingly intuitive problem of choosing the right word for the right context far from trivial even for native speakers of a language (Wang and Hirst, 2010).

Based on an analysis of a specific corpus of news articles on tsunamis, Bawcom (2010) maintains that word choice could be decided on from different perspectives. According to Bawcom (2010), word frequency is one factor that affects word choice between near synonyms as well as others such as register, style and purpose. Therefore, it would be very difficult for people to choose the most appropriate word from a group of synonyms with certain patterns in certain contexts.

The problem of differentiating near synonyms and choosing the appropriate lexis is especially daunting for second language learners (Mackay, 1980). The majority of vocabulary errors made by advanced language learners reflect learners' confusion among similar lexical items in the second language (Lee and Liu, 2009). Looking at a group of synonyms including *sheer, pure, complete* and *absolute*, Partington (1998) points out that:

> In reality, the choice of a lexical item is often extremely complicated. The learner/translator must know the collocational habits of the related items in order to achieve not just semantic feasibility, but also collocational appropriacy. (p. 39).

Recent years have also witnessed developments in exploring the use of corpus in teaching synonyms in ELT. For example, Wang and Wang (2005) conducted research on the word *cause* making use of CLEC (the Chinese Learner English Corpus) and BNC, and found that in the collocation of *cause* and *change*, *cause* and *great*(*er, est*), the Chinese learners overused the positive semantic prosody, and underused the negative semantic prosody. Wei (2006) investigates the words *commit, cause* and *effect*, based on CLEC, COBUILD and JDEST (Jiao Da English for Science and Technology). His analysis shows that compared with native speakers, Chinese EFL learners have a narrow range of collocations, vague semantic meanings, and underused or overused semantic prosody. He discusses the 'prosodic clash'

caused by the use of unusual collocations and explains from a functional perspective that native speakers create collocations to achieve particular effects—irony, insincerity and so on, while the Chinese learners' inappropriate use of collocations is a signal of pragmatic failure. Lu (2010) explores the collocational behaviour and semantic prosody of near synonyms through a corpus-based contrastive analysis between Chinese learners' English (CLE) and native English. The data show that near synonyms differ in their collocational behaviour and semantic prosody. CLE exhibits much deviation in both dimensions, and different types of CLE exhibit varying degrees of synonymous substitution and prosodic clash. The above CLE characteristics and developmental patterns were found to be closely related to word-for-word translation, and learners' inadequate knowledge of the collocational behaviour and semantic prosody of near synonyms was claimed to be the underlying factor. Pan (2010) makes a contrastive analysis of the collocational features of *cause* and *lead to* in SWECCL (Spoken and Written English Corpus of Chinese Learners) and BNC. The data show that English-major learners demonstrate similar semantic preferences to the native speakers, but that there are still great differences in their underlying collocational patterns.

Martin (1984) discusses instructional approaches to teaching synonyms and stresses the importance of providing students with common collocates. With the availability of computerized corpora, recent research has exploited concordances and collocation data for advising L2 learners in lexical choice (Yeh, et. al., 2007; Chang, et. al., 2008). Lee and Liu (2009) address the distinctions of synonyms in the context of second language learning. They conduct both corpus analysis and empirical evaluation to investigate the effects of collocation on near-synonym distinction. The result shows that collocation information may lead to learners' successful comprehension and use of synonyms. They also point out that the semantic differences between near synonyms and their implications are not easily recognized and are often not acquired by L2 learners. By providing a dynamic two-dimensional Near-Synonyms and Similar-Looking (NSSL) vocabulary learning system through WordNet, Sun et al. (2011) investigate whether matching exercises might increase Chinese EFL learners' awareness of NSSL words, particularly those that have the same translated meaning in Chinese, and they suggest that English teachers of Chinese students should spend more of their teaching time on distinguishing the exact meanings of these NSSL words. In addition, Danglli and Abazaj (2014) discuss the importance of lexical cohesion and word choice in the process of academic writing. They point out that language users need to be fully aware that selecting the right synonym in a given context requires knowledge of all the semantic dimensions of the word, which thesauruses alone often cannot give, and that correct use of synonyms can achieve accuracy as well as increase cohesion in a piece of writing.

3.7. Studies of near-synonyms in Chinese and from a cross-linguistic perspective

As McEnery and Wilson (1996) have pointed out, 'corpus linguistics is increasingly multilingual, with many languages and many varieties of those languages, being studied with the help of corpus data'. A couple of corpus studies have been conducted on Chinese synonyms.

Tsai, Huang and Chen (1996) present interesting work on differentiating a pair of near-synonyms, 高兴 (*gāo xìn*, happy, glad) and 快乐 (*kuài lè*, happy, joyful). By examining the correlation between their syntactic behaviours and lexical semantic properties, Tsai et al. show that syntactic constructs can be systematically explained in terms of two semantic features: <+control> and <+change-of-state>. Using the same methodology to find other semantic features that can predict syntactic patterns, Chief et al. (2000) examine the synonymous pair 方便 (*fāng biàn*) and 便利 (*biàn lì*), which both mean 'to be convenient', and they propose two semantic factors, namely beneficial role and lexical conceptual profile, to account for the differences of this synonymous pair in terms of their syntactic behaviours. For example, 便利 (*biàn lì*) cannot be modified by the negative marker 不 (*bù*)(not), because the profile of 方便 (*fāng biàn*) focuses on the whole positional event and can be negated like any proposition, while the profile of 便利 (*biàn lì*) focuses on the beneficial role rather than the whole sub-event. In order for the profile to focus on the beneficial role, the whole proposition must be presupposed and a presupposition cannot be negated/cancelled. In addition the semantics of 便利 (*biàn lì*) denotes a positive meaning and it would be semantically anomalous if the predicated were negated.

As part of a long-term project on the lexical semantic study of Mandarin verbs, Liu et al.'s (2000) work extends the research frontier to a new semantic field with four near-synonyms 投 (*tóu*), 掷 (*zhì*), 丢 (*diū*) and 扔 (*rēng*), all glossed as 'to throw'. To account for their semantic differences, two kinds of 'endpoints' are distinguished: the Path-endpoint (i.e., the Goal role) and the Event-endpoint (i.e., the resultative state). The analysis shows that although the verbs all describe a directional motion with a Path in their event structure, they differ in their participant roles and aspectual specifications. For example, 丢 (*diū*) may be used to describe the endpoint of an event, i.e., the resultative state of 丢 (*diū*), while 扔 (*rēng*) does not have a stative use.

Based on data from the Sinica Corpus, Huang & Hong (2005) investigate the differences between Chinese near synonymous sensation verbs and the sense distinctions provided by Chinese WordNet. As observed, the differences are shown by analyzing the lexical concepts and collocation distributions. Wu et al. (2011) investigate the collocational behaviours, semantic prosody, and morphological combinations of the two near synonymous verbs 帮忙 (*bāng máng*)(help) and 帮助 (*bāng zhù*)(aid). The study shows that the two near synonyms are normally not collocationally interchangeable, and that the semantic prosody is an important index to distinguish between 帮忙 (*bāng máng*) and 帮助 (*bāng*

zhù); to be specific, 帮助 (*bāng zhù*) takes many more negative collocates as compared to 帮忙 (*bāng máng*).

Huang and Hong (2005) analyse near synonyms in sensory verbs such as *see*, *touch* and *taste* in Mandarin Chinese and distinguish their lexical concepts, collocations, and core senses. On the other hand, Tsai (2010) examines the syntactic functions, occurrence frequency, and collocational relationship of 相同 (*xiāng tóng*)(the same), 一样 (*yí yàng*)(alike, the same) and 同样 (*tóng yàng*)(the same, similar) and compares their referential properties.

With the development of comparable corpora in English and Chinese, comparative/contrastive analyses have also been conducted. For example, Xiao and McEnery (2006) make a comparative empirical study of semantic prosody from a cross-linguistic perspective. The contrastive analysis shows that semantic prosody and semantic preferences are as observable in Chinese as they are in English. As the semantic prosodies of near synonyms and the semantic preferences indicated by their collocates are different, near synonyms are normally not interchangeable in either language.

To summarise, all the works discussed above have contributed valuably to our understanding of both English and Chinese synonyms, and by implication the methods of corpus linguistic studies are applied to Chinese synonymy. Nevertheless, these studies, valuable as being, contain some weaknesses. Firstly, most of the studies start with a pair or a small group of (usually three or four) putative synonyms and look at their differences; therefore their findings are local rather than generalising. Secondly there are no psychological studies on synonymy and very few studies have been conducted from a comparative perspective. To fill the gap, this thesis will adopt a corpus-driven approach to examining synonymy by looking at a large group of (over ten) possible synonyms. In addition, this study will explore the psychological aspect of synonymy and also conduct a comparative study, specifically between English and Chinese. Lexical Priming (Hoey, 2005) seems to be appropriate to serve as a theoretical framework and the next section gives a brief review of the theory of Lexical Priming.

3.8. Lexical priming and synonymy

The theory of Lexical Priming (LP) was proposed by Michael Hoey in 2005. Based on corpus analysis, LP gives explanations of the existence of important phenomena unearthed by corpus linguistics including collocation, colligation, and semantic association from a psychological perspective (discussed in the previous chapter).

The word 'Priming' is originally a psychological term, referring to 'an implicit memory effect in which exposure to one stimulus influences a response to another stimulus' (Wikipedia accessed at https://en.wikipedia.org/wiki/Priming_(psychology) on 30th April, 2016,). Based on psychological experimental developments and the corpus linguistic analysis of large amount of naturally occurring

data, Lexical Priming (Hoey, 2005) argues that vocabulary acquisition occurs in the process of repeatedly encountering words or phrases in different contexts. In other words, people are mentally primed with words through encounters in speech and writing and they become cumulatively loaded with the contexts and co-texts of the words or phrases in question in the process of encountering them. This process is the way people are 'primed' for language use and recognition/interpretation. Hoey (2005) makes an analogy between the mental concordance and the computer concordance and points out:

The computer corpus cannot tell us what primings are present for any language user, but it can indicate the kind of data a language user might encounter in the course of being primed. It may suggest the ways in which priming might occur and the kind of feature for which words or word sequences might be primed. (p. 14)

Lexical priming has made a number of claims. In particular, it claims that:

1. Every word is primed to occur with particular other words; these are its collocates.
2. Every word is primed to occur with particular semantic sets; these are its semantic associations.
3. Every word is primed to occur in association with particular pragmatic functions; these are its pragmatic associations.
4. Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.
5. Co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations.
6. When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.
7. Every word is primed for use in one or more grammatical roles; these are its grammatical categories.
8. Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations.
9. Every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations.
10. Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations.

(Hoey, 2005, p. 13)

It is the fifth claim which concerns synonymy that is most closely relevant to the current study (though others are also relevant). As the theory of Lexical Priming claims to apply to different languages, this study looks at both English and Chinese synonymy within the framework of Lexical Priming.

Chapter 4 The Psychological Reality of Synonymy

4.1. Introduction

The previous two chapters have reviewed the literature in both corpus linguistics and synonymy. It has been shown that the notion of synonymy has been taken for granted and the notion in traditional linguistics has been challenged or at least modified by the corpus approach. It seems that the validity of the notion of synonym also needs reconsideration. To test whether the concept has psychological validity, this chapter will report a psychological experiment to explore the psychological reality of synonymy. The purpose is to set up a preliminary stage for the later corpus analysis.

4.2. Different psychological status of synonymy and antonymy

Synonymy and antonymy, two common language phenomena that are often grouped together, seem to have a different status in both daily life and psychological/linguistic studies. Jones (2002) has pointed out that 'it has been widely documented that children tend to grasp the concept of oppositeness at a very early age, [and] together with other childhood learning exercises (such as counting, reciting nursery rhymes and distinguishing between colours), recognising antonyms seems to be a natural stage in an infant's linguistic development' (p. 1-3). Children often learn antonyms in pairs rather than as single items, for example *big vs. small*, *hot vs. cold*.

Antonymy is the 'most readily apprehended' (Cruse, 1986, p. 197) of sense relations and many examples become deeply ingrained in our mental lexicon from infancy. Clark (1970) has pointed out that even though word association testing also elicits synonyms and general collocates, informants tend to provide antonyms more often than anything else when they are asked to 'say the first thing that comes into your head'. Jones (2002) argues that

> '[I]t seems efficient to learn closely related words in tandem, yet it is difficult to think of other word pairs which are learnt in the same fashion as antonyms. One would not necessarily feel a similar urge to learn synonyms in unison, nor would one find it problematic to fully understand a superordinate term without first being taught all of its corresponding hyponyms' (p. 3).

Both daily life experience and research study findings seem to support the psychological reality of antonyms. The question then arises as to whether there is an equivalent psychological reality to synonyms. The concept of similar meaning seems unproblematic; words such as *big* and *large*, *cold* and *freezing* are comfortably recognised as having similar meanings by any native speaker of English. There is no doubt that people have a receptive understanding of synonyms. But the question is not whether people can recognize synonyms, but rather whether they can produce synonyms.

4.3. Purpose and the research questions

The experiment reported in this chapter is intended to explore the psychological reality of synonymy. To be specific, this chapter aims to answer the following three research questions:

(1) Do people have a sense of synonymy? In other words, do people have a sense of sameness in lexis?

(2) Do people share, or differ in, the meaning they ascribe to their sense of synonymy?

(3) If they differ in the way they produce synonyms, what might be the reasons for these differences? Is it a psychological difference?

4.4. Methodology: word association test

To test the psychological reality of synonyms, a word association test seems appropriate, in which subjects are given a list of prompt words and asked to give a response immediately, and it was this kind of test that led to the recognition of the psychological reality of antonyms.

Word association tests have been regularly utilised as an elicitation tool in the belief that word associations reflect fundamental characteristics of the relations between words in the mental lexicon (Nissen and Henriksen, 2006). They complement the evidence of intuition, and provide a wealth of data for which semantics must provide some explanation (Leech, 1981), because 'even the most preliminary analysis of the word-association game reveals its kinship with language comprehension and production' (Clark, 1970). Over many years, word-association tests carried out by psychologists have yielded much detailed information (Postman and Keppel, 1970), confirming the use by informants of relations between words such as synonymy, antonymy, hyponymy, etc., but none appears to have been conducted in order to investigate the psychological reality of synonymy.

Clark (1970) states that when people are presented with one word as a stimulus and asked to produce as a response 'the first word that comes into their head', there will be a fair degree of consistency in the results, provided that the responses are made without reflexion or hesitation. He claims this is because 'all speakers of a language have met the words with which they are familiar or at least the most common words in the same contexts' (p. 271).

For him, the response time is a very important parameter. He emphasises:

> When the player is allowed to take his time, he generally reacts with rich images, memories, or exotic verbal associations, and these give way to idiosyncratic, often personally revealing, one-word responses. **But when he is urged to respond quickly, his associations become more**

> **'superficial', less idiosyncratic, and more closely related in an obvious way to the stimulus; responses are much more predictable in that they are the ones almost everyone else gives to the stimulus.** (Clark, 1970, p. 272) (The sentences in bold are my own emphasis).

Therefore, if people have a shared sense of synonymy, the informants in word association test should provide synonyms with a high degree of consistency when they are asked to respond quickly.

4.4.1. choice of prompt words for the test

Before the test, the prompt words were chosen carefully. To avoid satisfying my own presuppositions, the prompt words were not subjectively chosen by me, but identified independently according to the following criteria. First, I used Google to search for the most commonly used synonyms in English and a website which lists synonyms for the 86 most commonly used words in English appeared in the top entries of the search result (http://justenglish.me/2014/04/18/synonyms-for-the-96-most-commonly-used-words-in-english/). Considering that the use of too many words as a prompt might cause participants to lose their focus on the test and that too few words might compromise the result, and keeping in mind the need to distribute the prompts across three lexical categories (adjectives, nouns and verbs) it was decided to select 30 prompt words. Adverbs and functional words were excluded in the current experiment but would be worth later exploration. An initial twenty-five words were decided on as the prompt words for the test on the criteria of choosing the top ten from each lexical category and also taking account of whether everybody would be equally familiar with the prompts. Table 4.1 lists the twenty-five words and their synonyms provided by the website.

amazing	incredible, unbelievable, improbable, fabulous, wonderful, fantastic, astonishing, astounding, extraordinary
brave	courageous, fearless, dauntless, intrepid, plucky, daring, heroic, valorous, audacious, bold, gallant, valiant, doughty, mettlesome
famous	well-known, renowned, celebrated, famed, eminent, illustrious, distinguished, noted, notorious
happy	pleased, contented, satisfied, delighted, elated, joyful, cheerful, ecstatic, jubilant, gay, tickled, gratified, glad, blissful, overjoyed
neat	clean, orderly, tidy, trim, dapper, natty, smart, elegant, well-organized, super, desirable, spruce, shipshape, well-kept, shapely
true	accurate, right, proper, precise, exact, valid, genuine, real, actual, trusty, steady, loyal, dependable, sincere, staunch
calm	quiet, peaceful, still, tranquil, mild, serene, smooth, composed, collected, unruffled, level-headed, unexcited, detached, aloof
fair	just, impartial, unbiased, objective, unprejudiced, honest
quiet	silent, still, soundless, mute, tranquil, peaceful, calm, restful
difference	disagreement, inequity, contrast, dissimilarity, incompatibility
idea	thought, concept, conception, notion, understanding, opinion, plan, view, belief
trouble	distress, anguish, anxiety, worry, wretchedness, pain, danger, peril, disaster, grief, misfortune, difficulty, concern, pains, inconvenience, exertion, effort
place	space, area, spot, plot, region, location, situation, position, residence, dwelling, set, site, station, status, state
story	tale, myth, legend, fable, yarn, account, narrative, chronicle, epic, sage, anecdote, record, memoir
begin	start, open, launch, initiate, commence, inaugurate, originate
cry	shout, yell, yowl, scream, roar, bellow, weep, wail, sob, bawl
decide	determine, settle, choose, resolve
describe	portray, characterize, picture, narrate, relate, recount, represent, report, record
explain	elaborate, clarify, define, interpret, justify, account for
help	aid, assist, support, encourage, back, wait on, attend, serve, relieve, succour, benefit, befriend, abet
plan	plot, scheme, design, draw, map, diagram, procedure, arrangement, intention, device, contrivance, method, way, blueprint
strange	odd, peculiar, unusual, unfamiliar, uncommon, queer, weird, outlandish, curious, unique, exclusive, irregular
fear	fright, dread, terror, alarm, dismay, anxiety, scare, awe, horror, panic, apprehension
answer	reply, respond, retort, acknowledge
look	gaze, see, glance, watch, survey, study, seek, search for, peek, peep, glimpse, stare, contemplate, examine, gape, ogle, scrutinize, inspect, leer, behold, observe, view, witness, perceive, spy, sight, discover, notice, recognize, peer, eye, gawk, peruse, explore

Table 4.1 Prompt words chosen from synonyms of most commonly used words in English online

As may have been noticed, some of the prompt words belong to more than one lexical category. A quick search of each word with different lexical categories in BNC was conducted. Table 4.2 shows the percentages of each lexical category to which the words belong. It can be seen that nine words are dominantly adjectives with a proportion of over 85% belonging to this category, four are dominantly nouns with over 75% and six are verbs with 75%.

	Adj.	N.	V.	Total
	Freq. (per million) (Percentage)	Freq. (per million) (Percentage)	Freq. (per million)(Percentage)	Freq. (per million)(Percentage)
amazing	1,822 (16.24)(99.9%)	2 (0.02) (1%)	/	1,824 (16.26)
brave	1,615 (14.40)(86%)	55 (0.49)(29%)	209 (1.86)(11.1%)	1,880(16.76)
famous	6,400 (57.05)(100%)	/	/	6,400 (57.05)
happy	11,340 (101.09)(100%)	/	/	11,340 (101.09)
neat	1,638 (14.60)(99.9%)	1 (0.01)(1%)	/	1,639 (14.61)
true	17,647 (157.31)(99.5%)	26 (0.23)(0.1%)	39 (0.35)(0.2%)	17,744 (158.20)
calm	1,111 (9.90)(34%)	892 (7.95)(27.3%)	1,262 (11.25)(38.7%)	3,265 (29.10)
fair	8,172 (72.85)(89.8%)	716 (6.38)(7.9%)	12 (0.11)(0.1%)	9,101 (81.13)
quiet	5,841 (52.07)(96.4%)	185 (1.65)(3.1%)	31 (0.28)(0.5%)	6,057 (53.99)
difference	/	18,897 (168.45)(99.9%)	6 (0.05)(0.1%)	18,907 (168.50)
idea	/	31,963 (284.90)(100%)	/	31,964 (284.90)
trouble	/	9,441 (84.16)(89.5%)	1,110 (9.89)(10.5%)	10,551 (94.05)
place	/	50,954 (454.20)(76.8%)	14,640 (130.50)(22.2%)	66,369 (591.60
story	/	17,878 (159.37)(99.9%)	1 (0.01)(0.1%)	17,879 (159.40)
begin	/	/	40,126 (357.70)(100%)	40,128 (357.70)
cry	/	2,145 (19.12)(27%)	5,792 (51.63)(73%)	7,938 (70.76)
decide	/	/	23,825 (212.38)(100%)	23,825 (212.40)
describe	/	/	23,376 (208.40)(100%)	23,376 (208.40)
explain	/	/	18,664 (166.37)(100%)	18,665 (166.40)
help	/	10,760 (95.92)(21%)	40,484 (360.90)(79%)	51,245 (456.80)
plan	/	21,707 (193.72)(62.2%)	13,187 (117.55)(37.8%)	34,926 (311.30)
strange	6,053 (53.96)(100%)	/	/	6,053 (53.96)
fear	/	9,006 (80.28)(62.2%)	5,117 (45.61)(35.3%)	14,478 (129.10)
answer	/	12,093 (107.80)(54.4%)	9,841 (87.72)(44.3%)	22,230 (198.16)
look	/	11,741 (104.66)(9.7%)	109,036 (972.00)(90.3%)	120,781 (1,076.70)

Table 4.2 Percentages of lexical categories associated with the Chosen words in BNC

The remaining five prompt words added to the list were *fruit*, *consequence*, *by-product*, *agree* and *accept*. These words will be analysed in Chapters 5 and 6 using the BNC corpus and one of the purpose of adding these words to the prompt list was to allow comparison of the results of the word association test with those of the corpus analysis. In addition, *fruit*, *consequence* and *by-product* are dominantly nouns and *agree* and *accept* are verbs in the BNC (See Table 4.3). Adding them to the final list enables there to be a balance in the three lexical categories. Finally, the metaphorical sense of *fruit* may be identified as a synonym of the word 'result'; it was therefore a matter of interest to see whether it prompted a different kind of response from the other 'result' prompts and whether we could find a possible link between synonymy and metaphor.

	Adj.	N.	V.	Total
	Freq. (per million)	Freq. (per million)(Percentage)	Freq. (per million)(Percentage)	Freq. (per million)Percentage)
fruit	/	4,989 (44.47)(99%)	50 (0.45)(1%)	5,040 (44.93)
consequence	/	7,763 (69.20)100%	/	7,763 (69.20)
by-product	/	254 (2.26)100%	/	254 (2.26)
agree	/	/	22,887 (204.00)(99.9%)	22,889 (204.00)
accept	/	/	19,841 (176.90)(99.9%)	19,843 (176.90)

Table 4.3 Percentages of lexical categories of additional words to the prompt list in BNC

The final prompt list therefore included thirty words altogether, all of which were content words. Three lexical categories (noun, verb and adjective) were included in the list, each represented by at least seven words with one dominant word class. The rest of the prompt words are either completely or dominantly distributed across two word classes. The word class of the items in the list was not given to the participants in the experiment because I also wanted to find out whether people store the synonyms according to word class.

4.4.2. subjects

Forty-two participants were involved in the word association test that I employed, of which nine were aged 16 or less, ten from age 17 to 25, thirteen from 26 to 40 and ten over 40 (Table 4.4). To reduce the variables in the experiment, I chose the adult participants from the same geographical and occupational background. All the participants were native speakers of English. The nine subjects aged 16 or less were from a local school near Liverpool. The adult participants studied/worked in Liverpool schools or universities. Before the test, participants were asked to fill in a form about their background. In addition to age and gender, their educational background and especially the subjects they studied at college/university were also elicited by the form.

Age Group	Under 16		17-25		26-40		Over 40		Total
Gender	Male	Female	Male	Female	Male	Female	Male	Female	
Number of participants	4	5	5	5	5	8	5	5	
Total	9		10		13		10		42

Table 4.4 Number of participants from different age groups and genders

4.4.3. test procedure

When the experiment was conducted, the task was first introduced to the participants. Then the following instructions were given: 'In this experiment, thirty words will be shown to you. For each word you will be given thirty seconds. Please write down as many synonyms as possible for each word.' In addition, to reduce anxiety of the participants and obtain a more accurate result, an explanation was given to the participants that it was not a test and the experiment was not interested in any individual's performance. Finally the informants were given thirty prompt words and asked to write down as many synonyms as possible within a limited time (30 seconds for each prompt word).

4.5. Result and discussion

In the following section, three research questions will be addressed.

4.5.1. sense of sameness in meaning

The first research question was: Do people have a sense of synonymy? In other words, do people have a sense of sameness of lexis?

Before answering this question, one point needs to be clarified, that is, knowing the term synonymy and having a sense of synonymy do not mean the same thing. If people have a sense of sameness, even if they don't know the term 'synonym', they should immediately understand the concept when the term is explained. Equally they may have no sense of sameness even if they have heard the term 'synonymy'.

Most of the participants immediately knew the term 'synonymy'. Only two out of forty-two asked what a synonym was. After being given an explanation and examples, no one seemed to have problems in understanding the concept. In other words, in spite of the fact that some did not know the term synonymy, in principle they had no difficulty in understanding the concept of sameness of lexis.

After the data were collected, all the words provided by the participants were examined. Table 4.5 gives examples of some prompt words and the synonyms provided. The number in bracket shows how many people (out of 42 participants) have offered that word as a synonym of the prompt. For example, *see* (23) means that the word *see* has been provided as a synonym of *look* by 23 (out of 42) participants in the association test. I underlined those words which I thought reasonable as synonyms (details discussed later).

The results showed that with some exceptions (to be discussed below), most of the words offered by my informants could be reasonably considered as synonyms of the prompt word. What needs to be noted here is that there is difference between identifying and producing synonyms. In some cases people have no problem in identifying synonyms immediately (for example *cold* and *freezing*), and in other

cases people may need time to decide whether words are synonyms or not (for example *result* and *by-product*). Both such situations test the receptive features of synonyms and receptive identification of synonymy is certainly one aspect of the psychological reality of synonyms. However whether people are capable of offering synonyms in response to prompts is another matter, testing productive identification. As has been mentioned before, Clark (1970) maintains that word association tests are good at eliciting a closely related stimulus and response when participants are asked to act fast. Therefore if the participants in the experiment were to provide predictable synonyms within a short period of time, it would be reasonable to say that there was a psychological reality to synonymy. On the other hand, if they were unable to provide predictable synonyms within a short period of time, we would have to conclude that the psychological reality of synonymy was limited to recognition.

As it happens, though, the results of the experiment are not open to such a simple interpretation. They in fact show us a very complicated picture. Take *famous* as an example. Thirty-three (79%) participants provided *well known* as a synonym, which indicates most people stored these two words as synonyms in their minds. About 38% (16 out of 42) of the participants offered *celebrity* as a candidate synonym. Due to the different grammatical categories of the two words it may be arguable whether they are synonyms or not, but it seems to indicate the closeness of the two words in some people's brains. On the other hand, *renowned*, a word I would consider to be a qualified synonym of *famous*, was produced by only 4 (less than 10%) participants. The reason might be the low frequency of *renowned* in daily use, but this also reflects a possible gap between perceptive and productive aspects of synonymy.

To provide more evidence on the gap between (me) judging/identifying synonyms and (the participants) providing synonyms, I looked at all the words provided by the participants for *famous* and underlined those which I thought reasonable as synonyms; these are given in Table 4.5. The underlined words include *well-known*, *known*, *renowned*, *recognised*, *celebrated* and *noted*. In addition *infamous* and *notorious* seem to be hyponyms, being particular ways of being famous (discussed in detail later). The words *celebrity* and *stardom* might be included as synonyms if we ignore grammatical category; *star* and *popular* are arguable. Finally those I would not consider as synonyms include *legend*, *icon*, *starring*, *rich*, *wag*, *film star*, *stardom*, *important*, *liked*, *familiar*, *recognisable*, *aware*, *remembered* and *starring*.

look	see (23), stare (14), observe (9), glance (9), view (8), gaze (5), watch (4), sight (3), focus on (4), peer (3), regard (3), search (3), glare (3), seek (2), perspective (2), peep (2), peek (2), browse (2), visualise (2), notice (2), check (2), style (2), fashion (2), scan (1), hunt (1), saw (1), stared (1), fixate (1), squint (1), pry (1), reflect (1), perceive (1), acknowledge (1), eye sight (1), gape (1), glimpse (1), appearance (1), pursue (1), consider (1), vision (1), examine (1), experience (1), oversee (1), seeing around (1)
famous	well known (33), celebrity (16), popular (15), known (7), star (6), renowned (4), infamous (3), notorious (2), legend (2), icon (2), recognised (1), celebrated (1), starring (1), rich (1), starry (1), wag (1), film star (1), stardom (1), important (1), liked (1), familiar (1), recognisable (1), aware (1), remembered (1), noted (1)

fear	scared (26), horror (7), frightened (7), anxiety (7), terror (7),terrified (6), scare (6),worried (5), fright (5), nervous (4), phobia (4), afraid (4), anxious (4), dread (3), worry (3), panic (2),petrified (2), frightful (2), discomfort (1), shock (1), timid (1), unsettlement (1), apprehensive (1), apprehension (1), alarmed (1), trepidation (1), fearful (1), stressed (1), petrified (1), threatful * (1), terrify (1), frighten (1), harm (1), nervousness (1), wary (1), unexpected (1), challenged (1), unexplained (1), unease (1), extreme (1), chilled (1), danger (1), panicked (1), unconfident (1)

Table 4.5 Examples of prompt words and their putative synonyms provided by participants

To sum up, although some cases need to be further discussed, based on the above results it is reasonable to say that people do have a sense of sameness in lexis and most of the time they are able to provide words with similar meaning when prompted to do so.

As just noted, it seems that the participants have provided a variety of words as candidate synonyms. The differences in synonyms provided by the participants may suggest different ways that synonyms are stored in our brains. When people identify that two words are synonyms, we cannot guarantee that the two words are stored closely together. On the other hand, if people produce the same/similar synonyms very quickly, these synonyms must have been stored somewhere close enough to each other in the brain that they can be recalled immediately. Therefore it seems that synonymy has a psychological reality, but it is different from and more complicated than that of antonymy. The next section will look at these complications in details.

4.5.2. variations in the candidate synonyms offered

My second research question was: Do people share, or differ in, the meaning they ascribe to their sense of synonymy? In other words, does synonymy mean the same or different thing to people?

If people share the same sense of synonymy, they should give the same, or at least a very similar, list of synonyms to the same prompt word. However, the result showed a different picture. For each word in the prompt list, a variety of words was offered by the informants.

Firstly, the synonyms provided by the participants are not identical to those provided on the website. Take *amazing* for example (Table 4.6). Nine synonyms are provided on the website, namely *incredible, fabulous, unbelievable, improbable, wonderful, fantastic, astonishing, astounding* and *extraordinary*. The test has however elicited a different set, altogether 40 putative synonyms, of which 25 are offered as synonyms by only one or two participants. In the website list, *improbable* and *astounding* are included as synonyms of *amazing*, but these do not appear in the test-elicited list at all. On the other hand, *brilliant, great, good, awesome* and *excellent* are at the top of the list elicited by the test, but do not appear on the website list. In table 4.6, the synonyms provided both by the website and the participants in the experiment are in italics and those only offered on the website but not by

the participants are in bold.

amazing	Synonyms from website	incredible, unbelievable, **improbable**, fabulous, wonderful, fantastic, astonishing, **astounding**, extraordinary
	Synonyms provided by participants	*fantastic* (24), brilliant (23), great (16), *wonderful* (13), *fabulous* (12), good (10), awesome (8), *incredible* (7), excellent (7), super (5), *astonishing* (4), *unbelievable* (4), superb (4), *extraordinary* (3), wow (3), brilliance (2), stunning (2), stupendous (2), cool (2), terrific (2), tremendous (2), perfect (2), smashing (1), powerful (1), exceptional (1), happy (1), startling (1), shocking (1), nice (1), magnificent (1), magical (1), delightful (1), exciting (1), unreal (1), formidable (1), special (1), beautiful (1), lovely (1), unique (1), spectacular (1)

Table 4.6 Comparison between synonyms provided by the website and the test participants

Secondly, for each prompt word, a large number of variations are provided as synonyms by the participants (Table 4.7). For example, the number for *amazing* in Table 4.7 is 40, which means that forty words have been provided by the participants in the experiment. What however needs to be mentioned here is that not all the words provided are considered by the author to be synonyms, though there is association of meaning between the words provided and words in query.

Prompt Word	Number of Synonyms provided	Prompt Word	Number of Synonyms provided	Prompt Word	Number of Synonyms provided
amazing	40	brave	32	famous	27
happy	44	neat	33	true	33
calm	43	fair	38	quiet	31
difference	32	idea	40	trouble	46
place	38	story	35	begin	25
cry	41	decide	33	describe	33
explain	35	help	31	plan	36
strange	39	fear	44	answer	27
look	45	fruit	24	consequence	30
by-product	28	agree	31	accept	36

Table 4.7 Number of putative synonyms offered by the participants for each prompt provided

Next, Table 4.8 shows the synonyms provided by the participants with the highest score for each prompt word, where the score refers to the number of participants who have provided the word as a putative synonym. The larger the number is, the greater the number of people who provided the word as synonym. For fifteen prompt words (half of the total), over 50% of participants have offered at

least one identical word as synonym (in bold in table 4.8). These pairs comprise *amazing* and *fantastic* (24), *brave* and *courageous* (24), *famous* and *well-known* (33), *neat* and *tidy* (40), *true* and *correct* (27), *quiet* and *silent* (28), *idea* and *thought* (31), *place* and *location* (25), *story* and *tale* (28), *begin* and *start* (42), *help* and *assist* (23), *fear* and *scare* (26), *look* and *see* (23), *consequence* and *result* (25), and *strange* and *weird* (34).

Prompt Word	Synonym with the Highest Score	Prompt Word	Synonym with the Highest Score	Prompt Word	Synonym with the Highest Score
amazing	**fantastic (24) (57%)**	brave	**courageous (24) (57%)**	famous	**well-known (33) (79%)**
happy	cheerful (13) (31%)	neat	**tidy (40) (95%)**	True	**correct (27) (64%)**
calm	peaceful (19) (45%)	fair	equal (13) (31%)	Quiet	**silent (28) (67%)**
difference	change (8) (19%)	idea	**thought (31) (74%)**	Trouble	naughty (9) (21%)
place	**location (25) (60%)**	story	**tale (28) (67%)**	Begin	**start (42) (100%)**
cry	sob (18) (43%)	decide	choose (20) (48%)	Describe	explain (13) (31%)
explain	describe (17) (40%)	help	**assist (23) (55%)**	Plan	organise/ze (12) (29%)
strange	**weird (34) (81%)**	fear	**scare (26) (62%)**	Answer	result (16) (38%)
look	**see (23) (55%)**	fruit	food (5) (12%)	consequence	**result (25) (60%)**
by-product	result (10) (24%)	agree	concur (14) (33%)	Accept	agree (17) (40%)

Table 4.8 Synonyms of highest score provided by participants

There are few cases where the lists of putative synonyms provided by different participants are the same, and these usually occurred when the informants only offered one or two synonyms. The results also show that the fewer putative synonyms that participants offer, the more likely it is that the lists will be the same. For instance, for the prompt word *famous*, seven people gave the same list of *well-known* and *popular*. Also for *neat*, eight persons offered the same list of *tidy* and *clean*. However in most cases, where more than three synonyms were provided, there are very few shared lists. For example, one subject from each age group was chosen randomly and his/her list for the prompt word *begin* was noted. The results are shown in Table 4.9.

Subject A (female, age under 16)	start, fresh, renew, create
Subject B (male, age 17-25)	start, initial
Subject C (female, age 26-40)	start, commence, firstly
Subject D (male, age over 40)	start, go, initiate

Table 4.9 Example of elicited synonym lists made by randomly chosen participates

This seems to indicate that there are few overlaps in the synonyms offered, which suggests that people do have different judgements on whether words are synonymous or not. In other words, people understand the concept and its border/boundary differently. A careful inspection of the candidate synonyms provided by the participants provides further evidence of this, as shown in the following section.

4.5.2.1. superordinate/subordinate and co-hyponym as candidate synonyms

Hyponymy refers to 'the lexical relation corresponding to the inclusion of one class in another' (Cruse 1986). For example, *jazz* is a hyponym of *music* since jazz is a type of music, and by the same token, *music* is a superordinate of *jazz*. Linguistic definition seems to distinguish hyponyms from synonyms very clearly. However in real language use these two concepts seem to be blurred.

First let's look at the prompt word *fruit* and some of its elicited words including *vegetables (2), orange* (2), *apple* (3), *pineapple* (1), *food* (5) and *snack* (1). *Vegetable* may be considered as co-hyponym of *fruit*. Usually *orange*, *apple* and *pineapple* would fall into the category of subordinates of *fruit* and *food* is superordinate of *fruit*. The case of s*nack* is complicated. It may have some evaluative sense as it refers to informal, small and casual meal. Depending on the culture, *fruit* may be or not considered as hyponym of *snack*.

Another situation seems to relate to co-reference. In an informal talk after the experiment one participant explained that if she had an apple for lunch, she could say 'I have some fruit for lunch', therefore *fruit* and *apple* refer to the same thing and that they can be synonyms. However, in this case she knew what she had for lunch and she was co-referring *apple* with *fruit*. It might imply that co-reference may be confused with synonymy.

Take *story* as another example. A number of words were provided as synonyms by the participants in the test, such as *tale* (28), *fable* (12), *narrative* (9), *fiction* (7), *novel* (4), *legend* (4), *anecdote* (4), *myth* (3) and *parable* (2). Some people may not have a problem in considering these words as all synonymous to *story*. However if we look at the definitions given to these words by Cambridge Dictionary Online, we could argue that, except for *narrative* being synonymous and *tale* perhaps being 'partially' synonymous with *story*, all the other words listed should be considered as hyponyms or subordinates of the word *story*.

narrative	a ***story*** or a description of a series of events
tale	a ***story*** especially one that might be invented or difficult to believe type
fable	a short ***story*** that tells a general truth or is only partly based on fact, or literature of this
fiction	the type of book or ***story*** that is written about imaginary characters and events and not based on real people and facts
novel	a long printed ***story*** about imaginary characters and events
myth	an ancient ***story*** or set of stories, especially explaining the early history of a group of people or about natural events and facts
legend	a very old ***story*** or set of stories from ancient times, or the stories, not always true, that people tell about a famous event or person
anecdote	a short, often funny ***story***, especially about something someone has done
parable	a short, simple ***story*** that teaches or explains an idea, especially a moral or religious idea
fairytale	a traditional ***story*** written for children that usually involves imaginary creatures and magic

(All the explanations are from Cambridge Dictionary Online. The words in bold are my own emphasis.)

In brief, the concepts of hyponymy and synonymy do not seem to be clearly distinguished from each other all the time and people may extend their notion of sameness in meaning and include hyponyms into the category of synonymy. This indicates that the notion of sameness in dictionary and in psychological reality may not be the same as each other. For some people, hyponyms or specific examples are also considered as having sameness or closeness of meaning.

4.5.2.2. metaphor, metonymy and meronymy

In the cognitive linguistic view, metaphor is defined as 'understanding one conceptual domain in terms of another conceptual domain' (Zoltan, 2010). Meronymy represents the relationship between a part and its corresponding whole. Metonymy is often regarded as 'a referential phenomenon where the name of a referent is used to stand for another referent' (Klaus-Uwe and Thornburg, 2003), e.g. *the crown* stands for *the monarchy*.

Even though *result* and *reward* were each offered as synonymous with *fruit* only once, they are still worth discussing, as *fruit* can be used metaphorically to mean 'the pleasant or successful result of work or actions' (sense offered by Cambridge Dictionary Online). The fact that these two words were provided as synonyms to *fruit* means that some people do consider metaphorical meaning when they seek for synonyms. The other possibility is that the metaphorical meaning has become fossilized for some people and they do not think they are being metaphorical at all.

Next, the prompt word and elicited word may be in a meronymous relationship. Examples are *fruit* and *seed* (3) as 'fruit' is the whole containing 'seed'. Finally *fruit* and *orchard* can be in a metonymous

relationship. Whether the last two pairs of words are synonymous may be controversial, but it suggests that for some people the notion of sameness is different from that of other people.

4.5.2.3. collocates as synonymous candidates

Interestingly participants also provided words which do not fall into any traditional category of semantic relations. For example, for the word *fruit*, participants offered *ripe* (1) and *exotic* (1) as well as *vegetables* (2), which may fall into the category of co-hyponymy (as mentioned before). A brief corpus analysis also shows that *vegetables* is the top collocate of *fruit* and usually appears in the structure *fruit and/or vegetables* in the BNC. In addition *ripe* and *exotic* also appear as collocates in the collocation list. Even though this does not constitute sameness of meaning between the prompt word and the elicited word, it indicates that the words are close to each other in the textual location and that this closeness may trigger the association of the word meanings.

4.5.2.4. candidate words which have textual primings

Hoey (2005) points out that in addition to collocation, semantic association, colligation and pragmatic association, lexis also has its textual primings, to be specific:

> 'Words (or nested combinations) may be primed positively or negatively to participate in cohesive chains of different and distinctive types (textual collocation).
>
> Words (or nested combinations) may be primed to occur (or to avoid occurring) in specific types of semantic relations, e.g. contrast, time sequence, exemplification (textual semantic association).
>
> Words (or nested combinations) may be primed to occur (or to avoid occurring) at the beginning or end of independently recognized discourse units, e.g. the sentence, the paragraph, the speech act turn (textual colligation).'
>
> (Hoey, 2005, p. 115)

There sometimes seems to be a causal relationship between the prompt word and the elicited word, for example, *fruit* and *healthy* (3). The word *healthy* is not a collocate of *fruit* or usually considered to be a synonym. Somehow the two words are associated or primed together in people's minds as there is a possible causal relationship between 'eating fruit' and 'being healthy'. Another example is *idea* and its elicited word *brainstorm* (3). These words may be related to each other in a cohesive chain, which Hoey

(2005) has labelled textual collocation.

An analysis of random 100 instances of *fruit* (as a lemma) in BNC seems to provide some evidence. Look at the following two examples:

Example 4.1:

Diets often fail in the long term because they are too demanding on will-power. In some cases they are also nutritionally unsound. And most diets are not flexible enough for you to indulge yourself occasionally. Rather than concentrate on restrictions it is much easier, at least initially, to consider the positive aspect of ***healthy eating***. Are you having enough ***fruit***, vegetables, low-fat milk, wholegrain bread, and cereals? Does your food supply you with enough calcium, iron, and vitamins? Are you having the right kinds of fats (polyunsaturated rather than hard, saturated fats)? It is not enough to rely on vitamin pills and hope for the best. A multi-vitamin and mineral tablet will not be enough to turn an unhealthy diet into a good one. You need to learn some basic facts about nutrition and the balance of different nutrients that you need at meals

Example 4.2:

Your general health will benefit from the following two points of the code, which may also reduce the risk of some cancers. Frequently eat fresh fruits and vegetables and cereals with high fibre content. Here is some evidence that foods rich in pro-vitamin A and vitamin C may give ***protection against cancer***. Most ***fruit*** and vegetables contain these vitamins and vitamin A is also present in fish. Food containing fibre may protect against cancer of the bowel. Fibre is found in fresh fruit and vegetables but mostly in wholegrain cereals and bread. These vitamins and fibre are best obtained through natural food.

The example 4.1 is an example of textual collocation, as '*healthy eating*' and '*having enough fruit*' are linked together in a cohesive chain. The example 4.2 seems to suggest textual semantic association, as 'foods rich in pro-vitamin A and vitamin C', 'protection against cancer' and 'most fruits and vegetables contain these vitamins' seem to be linked in a specific semantic relation.

There are other cases which seem to be related to textual colligation, for example, for the word *start,* informants provided *first/firstly, initial/initially,* and *introduction*. The reason why words *first/firstly, initial/initially* and *introduction* are provided as synonymous to *start* may be that they share similar textual primings. In other words, they are primed positively to participate in similar cohesive chains; occur in semantic relations of sequence, and appear at the beginning of recognized discourse units. To be specific, *first/firstly, initial/initially* are primed to occur at the beginning of sentences, paragraphs and introduction sections of texts. Similarly, the word *start* may also be primed to occur in the phrase

'*to start with*' and to occur at the beginning of sentences and paragraphs and in the introduction section of a text.

As discussed in the introduction (Chapter 1), people may not have difficulty in understanding the concept of noun, but they do not necessarily share the same sense of the concept; some may have a limited definition (e.g. solid objects such as *desk* and *car*) and some may extend the range it covers (e.g. *playing*). A similar situation applies to synonymy, that is, people have different sense of synonymy as some may include hyponyms, metaphors, meronyms, metonyms and other lexical relations, while others do not. The next part is to discuss what has caused these differences between participants.

4.5.3. causes for the differences in concept of synonymy amongst participants

The third research question was: If people differ in the way they use synonyms, what might be the reasons for these differences?

So far in this chapter, we have shown that synonymy is a psychological reality; in other words, people have a concept of sameness of meaning in lexis even though they may be unfamiliar with the term synonymy or synonyms. However, people do not have a shared sense of synonymy. For the same word, people may provide different synonyms. The present section is devoted to finding out whether these differences are caused by the prompt words used to elicit synonyms or by the people who have provided the candidate synonyms.

To answer this question, all the words offered by all participants as synonyms were summarised (see Appendix). Table 4.10 shows some examples.

Prompt	Elicited words
begin	start (42), go (8), commence (16), fresh (1), renew (1), create (1), make (1), first (4), firstly (3), outset (1), get going (2), introduce (1), first movement (1), introduction (1), open (1), off (1), initial (2), kickoff (2), proceed (1), embark (1), opening (1), birth (1), open (1), initiate (3), end (1)
fair	even (10), equal (13), same (1), balanced (8), sharing (1), king (1), helpful (1), both sided (1), agree (1), unbiased (3), just (10), open-minded (1), honest (5), true (2), pale (2), blond (4), proper (1), carnival (1), light (6), right (6), good (2), open (1), 50:50 (1), beautiful (1), correct (1), accurate (2), judge (1), mild (2), pretty (1), pleasing (1), fete (1), fairground (1), consistent (1), moral (2), accepted (1), reasonable (1), justified (1)
fruit	orange (2), apple (3), exotic (1), pineapple (1), vegetable (2), food (5), veg (2), healthy (3), vitamins (1), vegetarian (1), produce (4), result (1), reward (1), seed (3), bud (1), ripe (1), vegetation (1), pip (1), snack (1), natural (1), harvest (1), orchard (1), fresh (1), offspring (1)

Table 4.10 Examples of summarised elicited words

4.5.3.1. the relationship between candidate synonyms offered and types of prompt words

Based on the summary of elicited words, the prompt words were classified into three categories. In the first category, for the same prompt word there is one word with high consistency amongst the candidate synonyms provided. Examples are *begin, neat* and *strange*. For the word *begin*, all the participants (42/42) considered *start* as synonymous; 95% (40/42) of the participants wrote down *tidy* as a synonym of *neat*; and 81% (34/42) provided *weird* as a synonym of *strange*.

Word	Frequency in BNC	Standardised Frequency in BNC
start	48,690	434.00 per million
begin	40,128	357.70 per million
neat	1,639	14.61 per million
tidy	1,423	12.68 per million

Table 4.11 Frequency and standardised frequency of the selected word pairs

The similar frequencies of the pairs *begin* and *start,* and *neat* and *tidy* in the BNC seem to provide a possible explanation (Table 4.11). In her investigation of whether people choose the most frequently occurring synonym first when synonyms are available to describe the same event or situation in a text, Bawcom (2010) points out that though her hypothesis cannot be conclusively supported, the results of her analysis of a corpus of newspaper articles do suggest that synonyms used in cohesion are ordered with the most commonly occurring word first. Compared with other candidate synonyms, *begin* is the most frequent word among the synonymous candidates for *start*, and *tidy* for *neat*. Although Bawcom's point is centrally related to mine, my position is not the same as hers. According to Bawcom, there is a tendency for people to choose the most frequent word in describing the same event or situation. My point is that when people are asked to provide synonyms for a word, they usually go to the most frequently used synonym or the next most frequent word in the frequency list. From different perspectives, Bawcom and I are both arguing that frequency plays a vital (though not the only) role in eliciting and using synonyms.

With the pair *strange* and *weird,* there is a big difference in terms of the frequency in the BNC taken in its entirety. The frequency of *strange* is almost six times as high as that of *weird*. But less markedly different frequencies of the two words in the BNC spoken corpus seem to provide a possible explanation. Compared with other synonyms provided by the informants, *weird* is closest to *strange* in terms of standardised frequency in the BNC spoken corpus (see Table 4.12). This seems to support my previous claim that people tend to offer the most frequently used synonym or the next word in the frequency list. On the other hand it suggests that the mode (written or spoken) also plays an important part in eliciting synonyms.

What is also related to this point and needs to be mentioned here is the trend towards Americanisation and colloquialisation in spoken language due to the popularity of American mass media (see for example, Leech et al., 2009). *Weird* is an American word and frequently used in spoken language. The adoption of the American word might be one of the reasons why the word is offered as a synonym of *strange* by many people.

	Frequency (Standardized Frequency) in BNC	Frequency (Standardized Frequency) in Spoken BNC
strange	6,053 (53.96 per million)	437 (3.90 per million)
weird	1,056 (9.41 per million)	280 (2.50 per million)

Table 4.12 Frequency and standardised frequency of *strange* and *weird* in BNC and spoken BNC

The second category includes words which elicited more than one synonym with a similar frequency. For example: for the prompt word *amazing*, the words *fantastic* (with a frequency of 24) and *brilliant/brilliance* (23+2) were provided by the participants; and *equal* (13)*, even* (10) and *just* (10) were offered as synonymous to *fair*.

This category seems to be related to polysemy. When a word is polysemous and out of context, it frequently elicits several synonyms. Take *fair* as an example. Cambridge Dictionary Online lists over ten senses, of which two grammatical categories are offered, namely adjective and noun. As mentioned before, no grammatical category for the prompt words was provided to the participants, so it was expected that participants might provide synonyms from both grammatical categories. However, all the synonyms provided by the participants for this word were adjectives. A quick search in BNC shows that only 7.9% of instances of *fair* are used as nouns and that 89.8% are used as adjectives. The high percentage of adjective use of *fair* seems to indicate why people are primed this way.

The adjectives offered by participants as synonymous with *fair* were *equal* (13), *even* (10), *just* (10), *balanced* (8), *right* (6), *light* (6), *honest* (5), *reasonable* (1), *blond* (4) and *pale* (2). These elicited synonyms could be classified into two groups: first, words evaluating things or situations (altogether 53 occurrences). This group comprises *equal, even, just, balanced* and *right*. The second group contains words denoting colour or shade (in total 12 occurrences). This group comprises *light, blond* and *pale*.

In Chapter 3, I illustrated the problem with substitution/replacement being a criterion for synonymy. Here with the candidate synonyms provided by the participants in the word association test, it is not difficult to show that substitution/replacement as a criterion for words being synonymous only holds in some situations.

Example 4.3:

a. I am sure we can agree on a *fair* price. (76 hits of *fair price* in BNC)

b. I am sure we can agree on a *reasonable* price. (91 hits of *reasonable price* in BNC)

Example 4.4:

a. He does more than his *fair* share of housework. (235 hits of *fair share* in BNC)

b. He does more than his *equal* share of housework. (23 hits of *equal share* in BNC)

Example 4.5:

a. There is a *fair* chance it could be turned down. (59 hits of *fair chance* in BNC)

b. There is an *even* chance it could be turned down. (21 hits of *even chance* in BNC)

Example 4.6:

a. It would not be *fair* to Tony.

b. *It would not be *right* to Tony.

c. It is not *right* to treat Tony like that.

The words *fair/reasonable* in example 4.3 sentences a and b can be replaced by each other and the meanings remain much the same. A search in the BNC corpus shows 76 hits for *fair price* and 91 hits for *reasonable price*. In addition, 29 instances of *equal chance*, 3 of *reasonable share* and 11 of *even share* are found in BNC but none of *equal price*, *reasonable chance*, or *even price.*

In 4.4, a and b, *fair* and *equal* are interchangeable. The meanings of the two sentences however are slightly different. In 4.4.a, his *fair share* refers to 'decent share of work, or previously-agreed/accepted share of work' while *equal share* in 4.4.b refers to 'exactly the same amount of work'. In addition, the frequencies in BNC corpus are different, with 235 hits for *fair share* and only 23 for *equal share*.

However, in few situations could *fair* be substituted for *even, just, balanced* or *right* without changing the original meanings, even though *fair* share one or more collocations with some of the candidate synonyms. In example 4.5, *fair* and *even* share the same collocate *chance* and both words could be used in a sentence grammatically structured the same way. Nevertheless, the meanings are different as *'fair chance*' means 'quite *a high probability*' while '*even chance*' refers to a '50/50 chance'.

In most cases, the word in query in the sentence cannot be substituted with other words without changing the meaning. Paraphrase is the only option to maintain the meaning by using another linguistic structure. For example in 4.6, '*It would not be fair to Tony*' can be paraphrased as '*It is not right to treat Tony like that*'; because '*It would not be right to Tony'* could be understood as '*Tony would not think it is right*'. Even though there are very few cases in which *fair* and *right* are interchangeable (except *That's only fair*; *That's only right.*), they are still considered as synonymous by my informants.

The final category comprises words which seem to have elicited candidate synonyms with a lack of consistency. For example, for the prompt word *fruit*, participants provided various words including *orange* (2)*, apple* (3)*, exotic* (1), *pineapple* (1), *vegetable* (2), *food* (5), *veg* (2), *healthy* (3), *vitamins* (1), *vegetarian* (1), *produce* (4), *result* (1), *reward* (1), *seed* (3), *bud* (1), *ripe* (1), *vegetation* (1), *pip* (1), *snack* (1), *natural* (1), *harvest* (1), *orchard* (1), *fresh* (1) and *offspring* (1).

Three senses of *fruit* as noun and one as verb are listed by the Cambridge Dictionary Online:

1. *noun (PLANT PART)*

   the soft part containing seeds that is produced by a plant. Many types of fruit are sweet and can be eaten;
2. *noun (RESULT)*

   the pleasant or successful result of work or actions;
3. *(slang) a gay man. Many people consider this word offensive.*
4. *verb*

   When a plant fruits, it produces fruit.

As has been mentioned before, a number of words listed by the participants, somehow related to the first sense in the dictionary (CDO), would not normally be considered as synonymous but rather as superordinate/subordinate. These words are *apple* (2), *banana* (2), *pineapple* (1), *orange* (1) and *food* (3). After the experiment an informal talk were conducted and some participants were asked a few questions. One of the questions was 'why you think *apple* or *banana* is a synonym of *fruit*?' As mentioned before, one participant gave the reason that she could say 'I had some fruit for lunch' to mean 'I had an apple for lunch'. Some others said they had realised *apple* or *banana* were not synonymous with *fruit*, but within a limited response time, they could not think of any synonyms and could not help giving the first instinctive response. This situation, on the one hand, is related to the possible extended concept of synonymy for some people and on the other hand suggests that for some people close association in meaning does not always ensure synonymy.

In addition, *result* and *reward* each was included only once in the responses to *fruit* and *produce* four times. As shown above in the dictionary entry, *fruit* can have the sense of *result*. That only one person provided *result* or *reward* as synonyms for *fruit* is therefore surprising. It seems that most people do not

remember the metaphorical meaning of *fruit* when confronted with the word as a prompt.

Corpus analysis of *fruit* seems however to give us some hints. An analysis of a sample of 300 instances of *fruit* shows that 88.7% of instances are used with the first sense of soft produce of a plant, 10.3% used with the second sense (= the pleasant or successful result of work or actions) and 0.3% are used as verbs in BNC. The high frequency of the first sense and the relatively low frequency of the other senses in the corpus may be in line with the distribution of the synonyms offered in the experiment.

In summary, it seems that the type of prompt words may influence the responses of the participants. Some words may easily elicit the same response (e.g. *start* and *begin*); some may have multiple senses, therefore eliciting various responses (e.g. *fair* and *reasonable*, *equal*, *even*). For some words (e.g. *fruit*), it is too difficult for people to come up with synonyms, thus they offer words such as hyponyms, meronyms and even collocates, as candidate synonyms.

4.5.3.2. the relationship between candidate synonyms chosen and personal profile of participants

As mentioned previously, the association test also required the participants to provide their personal data, including age, gender and educational background. The next section explores whether choice or indeed awareness of synonymy varies according to age, gender and educational background.

4.5.3.2.1. age

The first step is to look at the possible links between synonyms and age. To begin with, the average number of synonyms provided by each age group was calculated and it was found that older participants tended to provide putative synonyms in a larger number. For the age group under 16, the average number of candidate synonyms provided is 1.88 per prompt. The average number increases with age, with 2.03, 2.51 and 3.12 candidate synonyms per prompt being provided respectively for age groups 17-25, 26-40 and over 40. Although the number of participants in this experiment is insufficient for us to draw a solid conclusion that the older informants are, the more synonyms they have provided, it does show a possible link between age and the number of synonyms provided.

Age Group	Under 16	17-25	26-40	Over 40
Average number	1.88	2.03	2.51	3.12

Table 4.13 Average numbers of candidate synonyms provided per prompt by different age groups

In addition to providing a greater number of putative synonyms, the older participants also provide a greater variety of putative synonyms. For example, participants under 16 listed the words *brilliant, extraordinary, good, fantastic, awesome, incredible, super, great, fabulous, wonderful* and *astonishing* as synonymous with the prompt word *amazing*. Age group 17-25 added *unbelievable, superb,* and *phenomenal* to the list, but left out *extraordinary, awesome* and *astonishing*. Participants of age 26-40 provided more words, namely *tremendous* and *stunning*, although again *astonishing* was missing. Finally, age group over 40 offered the words *spectacular, stupendous* and *exceptional* to the list, while *extraordinary* and *astonishing* were still left out.

The fact that older people tend to provide a greater number of putative synonyms with greater variety may relate to several issues.

Firstly, it may be due to a more flexible and richer interpretation of the concept of synonymy among adults. Hoey's lexical priming provides a possible explanation of the link between age and synonym storage. According to Hoey (2005), people are primed to use words in particular ways through various encounters in different contexts and co-texts. So priming is likely to be a cumulative process through various contexts over a long period of time. As Hoey (2005) points out that 'the priming of a word or word sequence is liable to shift in the course of an individual's lifetime, and if it does so, and to the extent that it does so, the word or word sequence shifts slightly in meaning and/or function for that individual' (p. 9). Therefore it is possible that older adults, compared to adolescents and young adults, may have formed a 'holistic' understanding of the concept of synonymy and also of the words serving as prompts.

Secondly, it may be relevant to the issue of education versus experience. Hoey (2005) states that every time we encounter a word, we either reinforce or weaken the primings of the word as the encounter may introduce the word either in a familiar or unfamiliar context or co-text, and therefore:

> '[P]riming is what happens to the individual and is the direct result of a set of unique, personal, unrepeatable and humanly charged experiences. Words come at us both as children and as adults from a plethora of sources. Parents, caretakers, friends, teachers, enemies, strangers (friendly and scary), broadcasters, newspapers, books, cards, letters, fellow pupils or colleagues – all at different times and to different degrees contribute to our primings. (p. 178)

As each individual has different experiences, 'cracks may occur as a result of conflict between a speaker's primings and someone else's primings'. One of the places where this is particularly likely to happen is in the educational system. Hoey (2005) states that 'explicit input from the teacher, in particular the correction of writing and, sometimes, speech in the classroom, often produces conflict with the primings achieved at home'(p. 180). In the current experiment all the informants were British

and received education in UK, but due to their difference in age and the different schools they went to, their education experiences will have varied to some extent.

Before the 1980s there was no national syllabus in UK and it is hard to find the English textbooks used during that period of time. However in the book The Complete Plain Words (first published in 1954, and second and third editions in 1973 and 1986), Gowers (1986) advocated that officials use simple and accurate words and avoid verbosity in their use of written English. For example, he suggests using simpler equivalents for compound prepositions such as *by means of (by, with, using), for the purpose of (to)* and *in the absence of (without)*. Furthermore, he also advises that 'if the choice is between two words that convey a writer's meaning equally well, one short and familiar and the other long and unusual, of course the short and familiar should be preferred' (p. 71). From the purpose of the book, which is to 'help officials in their written English as a tool of their trade', it may be guessed that at the time it was published people tended to use more complex synonymous words and that the successive editions of the book may have affected the way people used synonyms in both official and daily language.

In 1988 the Education Reform Act made considerable changes to the education system. The National Curriculum was introduced, which made it compulsory for schools to teach certain subjects and syllabuses. The 1988 national syllabus is no longer active on the National Curriculum website, so we are not certain about the situation of teaching synonymy in schools at that time. Nevertheless, in the 2013 National syllabus we noted that the term 'synonym' is included in the glossary for the English course for year 5 pupils.

Even though it falls outside the scope of this thesis to investigate how synonyms were/are taught by individual teachers in different classrooms, it is possible that the way teachers have taught synonyms at different periods has affected the way different age groups have responded in my experiment.

Next it may be noticed that even though on average older people provide putative synonyms in larger number with greater variety, there are still some words missing from the lists offered by the older age groups compared with the younger ones. One possible explanation for the words missing from the list is that some words have faded away as time passes by. As Eckert (1997) states, 'only the middle aged are seen as engaging in mature use, as 'doing' language rather than learning it or losing it'. On the other hand, the issue might be related to receptive and productive priming. According to Hoey (2005),

> Productive primings occur when a word or word sequence is repeatedly encountered in discourses and genres in which we are ourselves expected (or aspire) to participate and when the speakers or writers are those whom we like or wish to emulate. Receptive primings occur when a word or word sequence is encountered in contexts in which there is no probability, or even possibility, of our ever being an active participant – party political broadcasts, interviews with film stars, eighteenth-

century novels – or where the speaker or writer is someone we dislike or have no empathy with – drunken football supporters, racists, but also sometimes stern teachers and people of a different age group.

As this experiment was designed to ask informants to provide synonyms within a limited time, it may only elicit informants' productive primings with regard to synonyms, but not receptive synonyms. If they had been shown a long list of lexis and asked to choose from the list putative synonyms, it is possible that informants might have offered different results. This issue is worth exploration and recommended for further studies.

4.5.3.2.2. gender

As regards to gender, the average numbers of synonyms offered by male and female were calculated and it was found that females tended to provide more synonyms than males, with an average number per prompt of 2.455 and 2.325 respectively. For each age group, again females provided more synonyms than males, except for the age group over 40. The average number of synonyms per prompt for female and male is 2.05:1.68, 2.10:1.97 and 2.89:2.18 for age groups under 16, 17-25 and 26-40. However for age over 40 the number for female is 2.78 while it is 3.47 for male (see Table 4.14).

	under 16	17-25	26-40	over 40	Total
Female	2.05	2.10	2.89	2.78	2.455
Male	1.68	1.97	2.18	3.47	2.325

Table 4.14 Average numbers of candidate synonyms provided by different genders and age groups

An interesting analogy is with gender and colours. Research has shown women use a richer colour vocabulary than men. For example, DuBois (1939) found that women were more prompt than men in naming the 'right or more accurate' colour as women largely use elaborate colour vocabulary while men use basic colour words. Rich (1977) studied six groups subdivided by age and occupation and found that in describing colours, women, as compared to men, used 'more elaborate words' (for example, women use colour words such as *lavender, magenta* and *chartreuse,* while men use basic colour words: *red, orange, yellow, green, blue* and etc.) and tended not to repeat a colour word (a colour was described with another word by women but was described by men with exactly the same word as previously). Lakoff (1975) notes that women use a wider range of colour terms than men, and discriminate more precisely between different shades of the same colour. They use words such as *beige, ecru, aquamarine,* and *lavender* which are largely absent in the language of men.

A number of sociolinguistic studies have reported gender differences in language use and can provide insight into how men and women approach their social worlds. Within the social sciences, an increasing consensus of findings suggests that men, relative to women, tend to use language more for the instrumental purpose of conveying information; women are more likely to use verbal interaction for social purposes with verbal communication serving as an end in itself (e.g., Brownlow, Rosamon, & Parker, 2003; Colley et al., 2004; Herring, 1993). It is possible that women remember and tend to use more synonyms or 'elegant variation' for various social purposes. In different social settings, we have to use synonyms to achieve various purposes: either for establishing authority, posing professional status, or being polite or friendly, building rapport in a relationship. For example, Lakoff (1975) reports that men and women use a different set of adjectives to convey an opinion. As shown in the table below, Lakoff found some adjectives were only used by females and some were neutral with respect to gender. Although this research is now out of date and the social situation of men and women has changed, at least at some point in the past men and women may have differed in language use and this may be true with using synonyms.

neutral	women only
great, terrific, cool, neat	adorable, charming, sweet, lovely, divine

It should be emphasized that my concern is whether there is a link between gender and the results of my investigation into the psychological reality of synonymy, in other words, whether men and women remember and use synonyms differently. The following examples from other studies show support for the possibility.

Compliments, as social lubricates which 'create or maintain rapport' (Wolfson, 1983), are usually intended to make others feel good (Wierzbicka, 1987). Giving compliments is one of the common social behaviours. Based on a corpus of 484 naturally occurring compliments and compliment responses, Holmes (1986) analysed the distribution of compliments between New Zealand women and men. The result shows that women gave and received significantly more compliments than men did.

Examples 4.7 and 4.8 are from Holmes' corpus (1986) of naturally occurring compliment and compliment responses of New Zealand women and men. Example 4.7 is a dialogue between two female friends Sal and Meg. When Sal says Meg looks *terrific*, Meg responds to Sal's compliment with using the word *snazzy*, synonymous to *terrific*, achieving the purpose of giving her compliment back to Sal. In this way a rapport is established between the two females.

Example 4.7

Two women, good friends, meeting in the lift at their workplace.

SAL: hi how are you? You're looking just ***terrific***.

MEG: Thanks. I'm pretty ***good***. How are things with you? That's a ***snazzy*** scarf you're wearing.

Example 4.8 is a dialogue between two male colleagues Bill and Tom. Similarly, Bill gives his comments on Tom's appearance by saying '*you're looking very smart*'. Instead of giving complements back, Tom is embarrassed and explains why he dresses himself up.

Example 4.8

Two colleagues meet at coffee machine at work.

BILL: you're looking very smart today.

TOM: (Looking very embarrassed.) I'm meeting Mary and her mother for mother.

Again although the corpus was thirty years ago, it suggests that women used synonyms while men did not, which seems also to suggest that women produce more synonyms than men.

Examples 4.9 and 4.10 are from a project of Davis' (2003), which enquired into the relationship among talk, gender and learning. The English classroom activities were recorded in the north of England during the late 1990s. the purpose of the research was of course not on the topic of using synonyms; however as these were authentic language uses by school kids in the classroom, these findings of the analysis of these conversations instead of make-up (made-up) examples may be more reliable.

From examples 4.9 and 4.10, it can be seen that girls and boys show a big difference in language use in a classroom activity. The girls (in example 4.9) seem to use synonyms to support what others say and to establish a cooperative relationship while boys (example 4.10) rarely use synonyms to achieve that purpose. In example 4.9, the girl not only uses synonyms (*peaceful, clam, silent* and *relaxing*) but also co-hyponyms (e.g. *crops*, *barley* and *thyme*) to create rapport with what others say. However in example 4.10, the boys seem to be resistant to giving evaluative adjectives like *dazzling* and *gorgeous*.

Example 4.9:
CATH: Well it's got lots of field/ it's like countryside/
JULIE: Peaceful place/
LISA: Ermm(.)/
JULIE: it's got lots of flowers and (.)/
KATIE: Crops/
JULIE: It's got barley and thyme
LISA: Lots of fields and rivers
EMMA: Big countryside
LISA: it's got a river
JULIE: It's got wildlife
EMMA: Very *idealistic*
JULIE: Yes.
EMMA: *Like in a fairy tale*
LISA: Picturesque
JULIE: The mood is like *peaceful* and *silent* and nice and *relaxing*
EMMA: *Calm*
LISA: *Lazy*. *Laid back*.
JULIE: Yes.
EMMA: It seems as if .it's just got the scenery
JULIE: There's no like towns springing up everywhere
LISA: It's just fields and sky
JULIE: The same thing for everywhere for ever and ever.
LISA: The picture that is created is just like
EMMA: Fields that go on for ever and meet the horizon so it just looks like it's meeting the sky?
JULIE: Yeah
CATH: It's very *peaceful* picture.
JULIE: Yeah.

Example 4.10:
ANDY: What are we on?
PIERRE: Part three
KIRK: Oooh
PIERRE: The sun dazzling through the leaves like orange
KIRK: Pierre Pierre
PIERRE: and things it's gorgeous
KIRK: shut up/I'm not bothered
PIERRE: And the yellow gold
KIRK: you're just stupid you
PIERRE: And a GOLDEN GALAXY erm
KIRK: shut up Pierre
ANDY: Listen to him/ Listen to him/ oh God
KIRK: he'll shut uo now cos he's gonna smell it
ANDY: Oh God
KIRK: Oh God
PIERRE: Like crystals like with all colours coming out of it
KIRK: See? Do you HAVE to speak like that and moving your hands about like a queer?

4.5.3.2.3. subject field

Another factor worth considering in the results of the experiment is the possible effect of the subject/discipline of the participants. In terms of register, all the three dimensions of variation (field, mode and style) seem relevant to explaining why different synonyms were provided by the informants. Cruse (1986) defines 'field' as 'referring to the topic or field of discourse'. He explains 'there are lexical (grammatical) characteristics of, for instance, legal discourse, scientific discourse, advertising language, sales talks, political speeches, football commentaries, cooking receipts, and so on'. Obviously

profession and subject constitute typical foci of attention. Profession is not the only factor influencing 'field of discourse' (Cruse, 1986); people of different profession subjects however do get more access to their certain topic/field in certain style through certain mode.

The number of the participants with different professions/subjects in the experiment is insufficient to provide confident conclusions. The results however are suggestive. For example, a number of participants provided *naughty* as synonymous to *trouble* in the word association test. A look at the background information of the participants shows that they are all teachers from local schools. Apparently for them, 'trouble students' (though this is not a standard expression) means 'naughty students'. In addition, the only three participants who are studying medicine in universities provided *OCD* (*obsessive compulsive disorder*) as synonymous to *neat* in the experiment. We would predict that different subjects/disciplines give rise to different types of knowledge input, and therefore that people in different professions may be primed to use different lexis in different domains. What seems also relevant here is that due to difference in subject field some words which may be considered as synonyms by laymen are well distinguished by professionals. Although not derived from the results of the experiment, a perfect case in point is that linguists may distinguish 'learning' from 'acquisition', while the two words may mean the same for those who do not study linguistics.

Mode is concerned with 'the manner of transmission of a linguistic message' (Cruse, 1986). Whether the word is characteristically a written or spoken use seems to have on occasion influenced the elicitation of synonyms in my experiment. A number of teacher participants offered *go* as a synonym of *start*, and they explained they often gave such an oral instruction at the beginning of classroom activities.

Style refers to 'language characteristics which mark different relations between participants in a linguistic exchange' (Cruse 1986). As the experiment is designed to elicit candidate synonyms with a list of prompt words without context, we could not find out whether the words provided by the participants were affected by linguistic style; therefore this dimension will not be addressed here.

In brief, this section has explored possible causes for the differences in the concept of synonymy among the participants. Both the relationships of candidate synonyms offered with types of prompt words as well as with personal profile of participants were both discussed. However, some questions remained unanswered. For example, we have noticed that *result* and *reward* each was only offered once as synonyms of *fruit* in the experiment, and corpus data show that when used with a metaphorical meaning, *fruit* is usually in the singular form. Therefore, a question arises whether word form affects the elicitation of synonyms. Furthermore, in the case of words that are polysemous (e.g. *fair*), we cannot, because of the research design adopted, determine whether the lack of context may have made it difficult for a participant to decide which sense to respond to; the question of whether context might affect other

aspects of the performance of participants also remains unanswered, which is recommended for future study.

4.6. Conclusion

Synonymy and antonymy, although often discussed together, have a different status in psychological reality. Although research has shown that people tend to master antonyms at very early age, no studies have been conducted on the psychological reality of synonyms. The word association test reported in this chapter was designed to explore whether people share a sense of synonymy or not. Thirty prompt words were given to forty-two participants of four age groups drawn from a local school and from universities near Liverpool. Within a limited time, participants were asked to write down synonyms of the prompt words. The results show that the participants do indeed have a psychological sense of synonymy even though the terms 'synonyms' or 'synonymy' might be unfamiliar. In addition, participants were found to have differing concepts of synonymy with some working with a limited definition while others extended the concept to hyponymy, metonymy and meronymy. This chapter also has discussed the reasons why participants differed in their choice and range of synonyms and reported evidence that the differences were associated with the prompt words on one hand and with the different ages, genders and educational backgrounds of the participants on the other. As with some prompt words, most of people may give one identical synonym along with other variations. However, for some words, no single word was provided by all or most of the participants as the synonym of a particular prompt word and two (and sometimes more than two) words with similar frequencies were offered by participants. Finally for a small number of words, participants were found to provide no word with any consistency at all, but rather various competing putative synonyms.

The chapter went on to consider the possible link between age, gender, occupation and the storage of synonymy. It seems that the older the participants are, the larger number of putative variations they provide, which may be the result of priming by education and long years of reading experience. Furthermore, women were found to offer more synonyms as a response to prompts and it was suggested that this might be associated with the tendency noted in the literature for women to store and use more synonyms than men for various social purposes. Finally a link between synonyms offered as responses and the subject field of informants was explored.

To sum up, this chapter has shown that there is a psychological reality to synonymy, but it is not the same kind of psychological reality as that of antonymy. The results seem to suggest that due to the differences in prompt words and also in people's experiences, the concept of synonymy is not exactly the same in different people's minds. The next step is to see whether the corpus analysis of potentially

synonymy words could be consistent with or give explanations to the current finding. To be specific, next chapter will turn to the corpus approach to explore the potential uses of corpus linguistics for describing synonymy as well as discussing approaches to recognizing and differentiating synonyms.

Chapter 5 Corpus Approach to Notion of Synonymy

5.1. Introduction to the Chapter

The previous chapter has explored the psychological aspect of synonymy and the results obtained in the psycholinguistic experiment have shown that people may provide variations of candidate words as synonyms to the prompt words. These variations may be conventionally considered as superordinate/subordinate, co-hyponym, metaphor, metonym and even antonym, which indicates that the boundaries between synonymy and other lexical relations may not be as clear-cut as we thought.

This chapter will turn to the corpus approach to explore the potential uses of corpus linguistics for describing synonymy as well as discussing approaches to recognising and differentiating synonyms. The purpose is to check whether the analysis of corpus data supports the findings obtained in the psycholinguistic experiment reported in the previous chapter.

Among the approaches used in the recognition of synonyms, substitution has been one of the most persistent criteria (Palmer, 1981; Lyons, 1981; Cruse, 1986; Hoey, 1991; Sinclair, 1991; Stubbs, 2001). Traditionally, two words are considered synonymous in a sentence or linguistic context if the substitution of one for the other does not alter the truth value of the sentence. This explanation has however been shown to be not only ambiguous but impractical in determining whether candidate words are synonyms or not (see Chapter 3). Likewise componential analysis has also proved ineffective in defining synonymy and discriminating between synonymy and co-hyponymy (see Chapter 3 for details). Despite the deficiency of substitution and componential analysis in differentiating synonymy, no other approaches were proposed until the development of corpus linguistics. Since then, although a number of corpus investigations have been conducted into synonymy by looking at their collocations and semantic prosodies, there have been few holistic or systematic studies of synonyms.

As mentioned in the Introduction (section 1.5), this thesis is concerned to answer six overall research questions and this chapter will focus on the second one:

> If we find that synonymy has psychological reality, does the analysis of corpus data help to explain the findings obtained in psycholinguistic experiments?

This chapter will start with a corpus-driven analysis of potentially synonymous items to explore whether these words are really synonyms. By identifying the strength of similarities among the candidate synonyms, the study will explore whether the analysis of authentic language data justifies the retention of the concept of synonymy.

Lexical priming, a corpus-driven linguistic theory, offers an excellent explanation from a psychological perspective for the existence of key concepts in corpus linguistics such as collocation, semantic association, colligation and pragmatic association. It claims that people are primed to use words and phrases in particular ways through their encounters with these words and phrases in different contexts and co-texts. Drawing an analogy between mental concordances and computational concordances, Hoey (2005, p. 13) hypothesises that 'every word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word'. More specifically, every word is primed differently in terms of its collocations, semantic associations, pragmatic associations and colligations. Based on an analysis of hyponyms of SKILLED ROLE OR OCCUPATION, Hoey concludes that the collocational and colligational behaviours of the co-hyponyms are too variable to routinely allow generalisations about the set as a whole. On the other hand, his analysis of *result* and *consequence* shows that there are indeed shared primings for these synonyms though they also differ in the strength of their shared collocations, colligations, semantic associations and pragmatic associations. If these claims hold true of other sets of hyponymous and synonymous pairs, they may provide us with a potential approach for distinguishing synonyms from co-hyponyms.

This chapter seeks to investigate whether it is possible to define or describe synonyms and distinguish synonymy from other semantic relations such as co-hyponymy, meronymy and metaphor by looking at the collocational and colligational behaviours of a set of lexical items of closely related meaning. By examining a group of nouns within the framework of lexical priming, the study explores how a corpus analysis of candidate synonyms might help to sort out synonyms from other semantic relations. To be specific, this chapter is to explore the shared features of these candidate synonyms in terms of their collocations, semantic associations and colligations. The result of analysing a group of potentially synonymous words leads to the suggestion that the term 'synonymy' is an ineffective and simplistic term for such a complex language phenomenon. Despite this, lexical priming allows us to make progress in identifying behaviours of synonymy.

5.2. A corpus-driven analysis of potentially synonymous items

5.2.1. purpose and specific research questions of the chapter

As discussed in Chapter 3, semantic relations including synonymy, co-hyponymy, metonymy and metaphor seem to be commonly-used terms in linguistics; as we have seen, however, the distinction between them can be very tricky sometimes. Partington (1998) points out that 'although it would not be possible to examine all the contextual relations of a pair of items, by utilizing corpus data, it is

possible to examine large numbers of their co-textual relations, particular their collocational patterns' (p. 32).

Even though the terms for the different semantic relations are not unfamiliar to many linguists, and the boundaries between them are recognized to be blurred at times, they have not been much investigated with a view to finding ways of distinguishing them. Hoey's (2005) work might be one exception. Looking at a group of co-hyponyms such as *carpenter, architect, actress* and *accountant,* Hoey (2005) points out that the various hyponyms of SKILLED ROLE OR OCCUPATION are typically primed quite differently from each other, at least as far as collocation is concerned. However, he also suggests that the existence of characteristically shared primings will provide the conditions for a trustworthy definition of synonymy.

If this claim holds, the research questions of this chapter then are: 1. Can we sort out synonyms from co-hyponyms or words in other semantic relations via data-driven analysis? 2. How confident can we be that pairs or groups of words are synonyms, co-hyponyms or in any other semantic relations? 3. Can the results of corpus analysis help explain the findings in the psycholinguistic experiment with respect to different senses of synonymy among people?

5.2.2. Methodology

Hoey's analyses of synonyms and co-hyponyms led me to hypothesise that bottom-up analyses of lexical items might suggest ways of sorting out synonymy from other semantic relations in terms of their collocations, semantic associations and colligations. To test this hypothesis, a number of nouns which are potentially synonymous comprising *RESULT, OUTCOME, AFTERMATH, UPSHOT, SEQUEL, EFFECT, END-RPODUCT, BY-PRODUCT, FRUIT, IMPACT* and *CONSEQUENCE* (in capitalisation to refer to the lemma of the word) were chosen for corpus linguistic analysis. The list for analysis was created by reference to dictionaries and thesauri, in which previous linguists and lexicographers have offered their intuitions and introspections with regard to the candidate synonyms. These words were chosen for the following reasons. First, a number of corpus analyses have been done with some of the words in the group, for example, in Hoey (2005), and Xiao and McEnery (2006). Second, it is intended that the findings of the corpus analysis should be compared with the findings of the psychological experiment reported in Chapter 4, and it was therefore important that the words chosen should overlap with the items used in that experiment. Third, some of the words can be used as discourse markers, for example, *RESULT* and *CONSEQUENCE* can appear in the phrases *as a result* and *as a consequence*, which signal a discourse relation of cause and effect. By looking at these words, I am therefore not only analysing synonyms of individual words but also exploring

synonyms in discourse. And fourth, previous studies on synonyms have mainly focused on synonymous pairs, few on groups of four or five words, and none appears to have been conducted with a group of ten candidate words.

To answer the research questions of this chapter, the British National Corpus is analysed using Sketch Engine (Kilgarriff, 2003), a language analysis tool which offers various applications such as *word sketch* and *sketch difference* in addition to the expected *concordance* and *word list*. A word sketch is a one-page summary of the word's grammatical and collocational behaviour. It shows the word's collocates categorised by grammatical relations such as words that serve as an object of the verb, words that serve as a subject of the verb, words that modify the word etc. Word sketch difference is used to compare and contrast two words by analysing their collocations and by displaying the collocates divided into categories based on grammatical relations (Kilgarriff, 2003).

5.2.3. Results and discussion

5.2.3.1. frequency

5.2.3.1.1 raw frequency and standardised frequency in BNC corpus

The frequency of all the lemmas in the corpus is investigated first. The Sketch Engine offers the opportunity for different inquiries and a sample snapshot of a search entry (*RESULT* as an example) can be seen as follows:

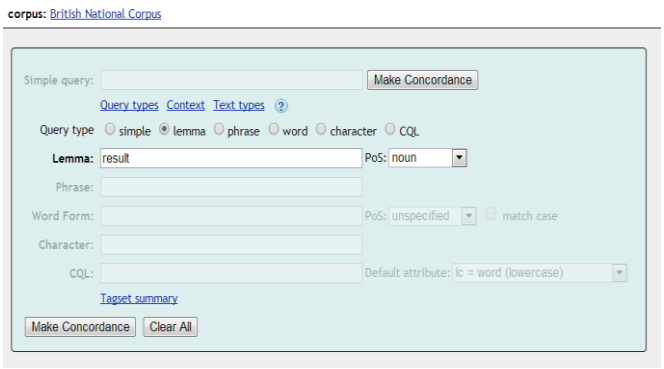


Figure 5.1 Snapshot of search entry in the Sketch Engine (take *RESULT* as an example)

All the query words are listed in order in terms of ranks of standardised frequency of the lemmas in the BNC (Table 5.1). Interestingly, these words could be paired with respect to their normalised frequency, for example *EFFECT* and *RESULT* with 296/297 per million, *IMPACT* and

CONSEQUENCE with 68/69 per million, *FRUIT* and *OUTCOME* with 41/45 per million, and *SEQUEL* and *BY-PRODUCT* with 2.3/2.7 per million.

Rank	Lemma	Raw frequency	Standardised frequency (per million)
1	*RESULT*	33,890	301.80
2	*EFFECT*	33,231	295.94
3	*CONSEQUENCE*	7,733	68.87
4	*IMPACT*	7,482	66.63
5	*FRUIT*	4,760	42.39
6	*OUTCOME*	4,524	40.29
7	*AFTERMATH*	685	6.10
8	*SEQUEL*	286	2.55
9	*BY-PRODUCT*	253	2.25
10	*UPSHOT*	154	1.37
11	*END-PRODUCT*	59	0.53

Table 5.1 Raw and standardised frequency of the lemmas in the BNC

However, this does not necessarily mean any pair can be considered to be synonyms. But if we are primed with words through encounters in various contexts, the frequency may suggest to some extent the order of words in the scale of synonymy; for example, *CONSEQUENCE* may be offered before *OUTCOME* as synonymous to *RESULT*. As Bawcom (2010) has pointed out, frequency is one of the most important factors in choosing synonyms. The frequency list may indicate one factor in how synonyms are stored in people's minds, and Chapter 4 has reported a small-scale experiment to test the storage of synonyms in people's minds. The result of the experiment will be compared with the frequency list in detail in this chapter.

5.2.3.1.2. frequency and word form in BNC corpus

Some studies have revealed that different word forms of a lemma behave differently and may denote different meanings. For example, Sinclair (1991b) shows the non-equivalence of singular and plural form of nouns (*eye* and *eyes*), and states that 'there is hardly any common environment' between the two word forms and that they 'do not normally have the capacity to replace each other' (p. 489). Therefore, I also checked the frequency of singular and plural forms of each noun in the BNC corpus

(Table 5.2). The result shows that different word forms of each lemma have different distributions in the corpus. The nouns whose ratio between singular and plural forms is over 8:2 include *IMPACT*, *FRUIT*, *OUTCOME*, *AFTERMATH*, *SEQUEL* and *UPSHOT*. How the differences of word forms affect the meaning of these nouns in the context will be explored in the latter part of this chapter.

Rank	Lemma	Word form	Raw frequency	Standardised frequency (per million)	Percentage
1	*RESULT*	result	19,040	169.60	56.2%
		results	14,847	132.20	43.8%
2	*EFFECT*	effect	22,606	201.30	68%
		effects	10,620	94.60	32%
3	*CONSEQUENCE*	consequence	3,390	30.20	43.8%
		consequences	4,343	38.70	56.2%
4	*IMPACT*	impact	7,230	64.40	**96.6%**
		impacts	251	2.24	3.4%
5	*FRUIT*	fruit	3,824	34.10	**80.3%**
		fruits	933	8.31	19.6%
6	*OUTCOME*	outcome	3,627	32.30	**80.2%**
		outcomes	897	7.99	19.8%
7	*AFTERMATH*	aftermath	682	6.10	**99.6%**
		aftermaths	3	0.03	0.4%
8	*SEQUEL*	sequel	247	2.20	**86.4%**
		sequels	38	0.34	13.3%
9	*BY-PRODUCT*	by-product	174	1.55	68.8%
		by-products	79	0.70	31.2%
10	*UPSHOT*	upshot	153	1.36	**99.4%**
		upshots	1	0.01	0.6%
11	*END-PRODUCT*	end-product	43	0.38	72.9%
		end-products	16	0.14	27.1%

Table 5.2 Frequency of singular and plural forms of each noun in the BNC corpus

5.2.3.1.3. frequency and text types

Hoey (2005) points out that primings are 'domain-specific'. Table 5.3 shows the occurrence of each lemma in different text types in matrix form. Five text types are categorized in BNC, namely written books and periodicals, written miscellaneous, spoken context-governed, written-to-be-spoken and

spoken demographic. In each cell of the matrix, two numbers are provided. The first one is the frequency of the lemma in the particular text type; and the second one is which is termed as 'relative text type frequency' (Rel for short, figure shown in percentage). The number is the relative frequency of the query result divided by the relative size of the particular text type. The number grows with higher frequency and gets smaller the greater the size of the text type. It can be interpreted as 'how frequent is the result of the query in this text type in comparison to the whole corpus'. For example, 'test' has 2000 hits in the corpus and 400 of them are in the text type 'Spoken'. Text type 'Spoken' represents 10 % of the corpus. Then the Relative Text Type frequency will be (400 / 2000) / 0.1 = 200 % and it means "test" is twice as common in 'Spoken' than in the corpus as a whole. (https://www.sketchengine.co.uk/rel/)

In Table 5.3 we can see that the words in query have different distributions in different text types.

	Lemma	Written books and periodicals	Written miscellaneous	Spoken context-governed	Written-to-be-spoken	Spoken demographic
		Raw frequency (Rel %)	Raw frequency (Rel %)	Raw frequency (Rel %)	Raw frequency (Rel %)	Raw frequency (Rel %)
1	*RESULT*	27,866 (104.10)	3,832 (152.70)	1,105 (53.10)	300 (67.50)	103 (7.4)
2	*EFFECT*	29,433 (109.60)	2,454 (97.50)	1,157 (55.40)	180 (40.30)	91 (6.5)
3	*CONSEQUENCE*	6,852 (109.50)	621 (105.90)	244 (50.20)	39 (37.50)	7 (2.10)
4	*IMPACT*	6,068 (98.80)	1,092 (189.70)	386 (80.90)	57 (55.90)	16 (5.0)
5	*FRUIT*	4,214 (104.80)	350 (92.90)	134 (42.90)	57 (85.30)	234 (111.50)
6	*OUTCOME*	3,898 (105.00)	522 (150.10)	130 (45.10)	39 (63.30)	15 (7.70)
7	*AFTERMATH*	640 (115.00)	38 (72.90)	4 (9.30)	7 (75.80)	1 (3.40)
8	*SEQUEL*	270 (111.20)	19 (83.50)	3 (15.90)	6 (148.80)	3 (23.70)
9	*BY-PRODUCT*	230 (112.30)	18 (93.80)	4 (25.10)	2 (58.60)	/
10	*UPSHOT*	139 (111.90)	7 (60.20)	8 (82.90)	/	/
11	*END-PRODUCT*	56 (121.80)	1 (23.20)	/	/	/

Table 5.3 Frequency and relative text type frequency of each lemma

It can be seen that there are 3,898 instances of *OUTCOME* in written books and periodicals, 522 instances in written miscellaneous, 130 in spoken context-governed, 39 in written-to-be-spoken and

only 15 instances in spoken demographic. This contrasts markedly with *FRUIT*, which occurs 234 times in spoken demographic as opposed to only 350 times in written miscellaneous (in relative frequency %, a difference between 92.9 and 150.5). This suggests that apparent synonyms do not distribute in the same ways across domains, modes and genres. The main value of the step, though, is that it provides us with both a statistical and methodological basis for the following analysis and discussion.

5.2.3.2. Collocation

The next step of the analysis concerns collocation. Firth (1957) states 'you shall know a word by the company it keeps', so the working hypothesis here is that overlap in collocation may reveal which words have meanings or senses that are closer to each other than others. On the other hand differences in collocation may also indicate divergence among synonyms.

I used Word Sketch to compare the collocations of each lemma. Word Sketch, one of the built-in applications in the Sketch Engine, is useful in providing a one-page summary lexical and grammatical description of the word in query. It shows the word's collocates categorised by grammatical relations such as words that serve as an object of the verb, words that serve as a subject of the verb, words that modify the word etc. The statistics used in word sketch is that of logDice, whose score 'has a reasonable interpretation, scales well on a different corpus size, is stable on subcorpora, and the values are in reasonable range' (Rychlý, 2008, p. 9). A sample snapshot of analysis result (taking *team* as an example) can be seen in figure 5.2.

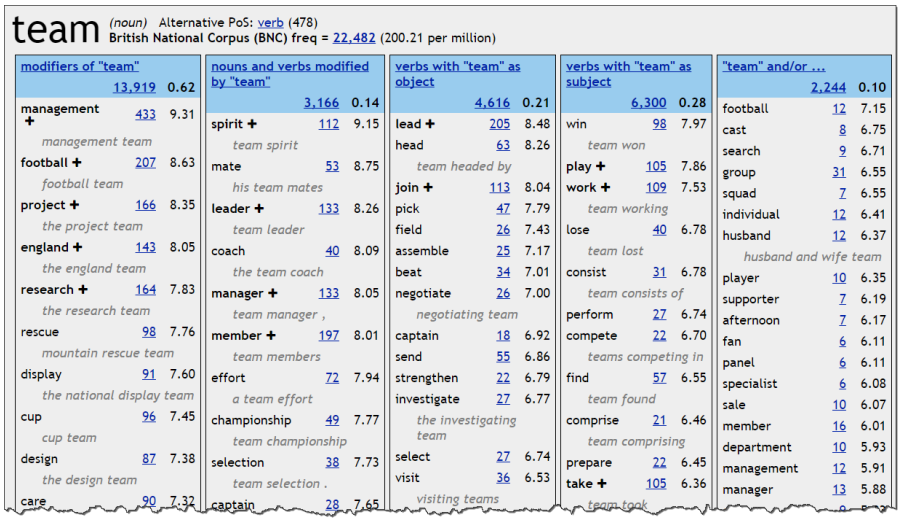


Figure 5.2 Snapshot of analysis result of *team* (as an example) with Word Sketch

Each lemma was analysed in Word Sketch and a large number of detailed results were elicited. The following sections will demonstrate the results along with the discussions one by one.

5.2.3.2.1 modifiers of the words in query

Table 5.4 lists all the collocates that function as modifiers to our query words in Word Sketch analysis. These collocates are generated on the basis of the collocational strength measured statistically. Collocation is bidirectional and collocation studies rely on some measurement of association. Raw frequency of co-occurrence can be misleading because if one item in the collocation is extremely frequent, then relatively high co-occurrence may just be the result of the overall high frequency of the item. Corpus linguistics has developed a number of calculations to determine relative degree of association, especially between individual words. Commonly-used measurements of lexical association include the mutual information (MI) score, the z-score, the t-score and the log-likelihood (Krenn and Stefan, 2001; Pearce, 2002; Ramisch et al., 2008). Both on-line and stand-alone language analysis tools adopt one or the other of these, sometimes in combination. MI overestimates the importance of collocations of low frequency, while t-score overestimates those of high frequency (Hamilton et al., 2007; Manning & Schütze, 2001). The measurement used in Sketch Engine is logDice, which has been argued to be more reliable since it is not biased by either too high or too low a frequency of the items in query (Kilgarriff and Kosem, 2013).

In Table 5.4, along with each collocates as modifiers, two figures are provided in the brackets, first being the frequency of the collocate and the second being the significance of the collocational association between that collocate and the query word.

Rank	Lemma	Collocates as modifiers (Frequency; Significance)
1	*RESULT*	end (270; 9.48), election (225; 9.21), direct (240; 8.78), examination (133; 8.68), test (146; 8.47), positive (152; 8.34), similar (166; 7.94), net (100; 7.92), final (154; 7.87), good (577; 7.81), exam (63; 7.69), preliminary (68; 7.64), experimental (66; 7.64), inevitable (61; 7.52), negative (67; 7.42), overall (76; 7.31), excellent (70; 7.29), disappointing (46; 7.23), poor (89; 7.22), satisfactory (45; 7.08), disastrous (41; 7.02), recognition (33; 6.68), interim (35; 6.66), same (177; 6.66), research (58; 6.63)
2	*EFFECT*	adverse (335; 9.38), side (358; 9.31), greenhouse (265; 9.09), profound (168; 8.38), beneficial (160; 8.35), immediate (189; 8.22), cumulative (141; 8.17), significant (213; 8.16), direct (200; 8.11), dramatic (145; 8.02), devastating (120; 7.94), knock-on (116; 7.94), overall (157; 7.91), long-term (131; 7.81), damaging (106; 7.76), possible (157; 7.70), opposite (111; 7.67), detrimental (95; 7.64), harmful (86; 7.48), indirect (89; 7.45), net (102; 7.44), combined (89; 7.44), negative (97; 7.42), positive (108; 7.40), sound (87; 7.33)
3	*CONSEQUENCE*	inevitable (88; 9.37), disastrous (62; 9.04), unintended (42; 8.71), adverse (51; 8.62), dire (42; 8.62), serious (125; 8.42), far-reaching (36; 8.40), likely (52; 8.36), damaging (37; 8.33), unfortunate (30; 7.92), logical (34; 7.84), direct (77; 7.78), tragic (25; 7.66), profound (25; 7.58), possible (71; 7.53), negative (34; 7.46), important (114; 7.40), practical (44; 7.29), harmful (17; 7.29), long-term (31; 7.20), fatal (18; 7.18), normative (13; 6.96), environmental (36; 6.95), unforeseen (12; 6.86), undesirable (12; 6.82)

4	*IMPACT*	environmental (176; 9.23), significant (138; 8.54), likely (54; 8.40), adverse (35; 8.05), immediate (63; 7.91), profound (30; 7.82), visual (41; 7.78), devastating (20; 7.43), major (133; 7.40), dramatic (31; 7.37), negative (31; 7.31), direct (53; 7.24), lasting (18; 7.23), differential (17; 7.20), potential (39; 7.09), considerable (49; 7.07), maximum (26; 7.05), tremendous (19; 6.96), enormous (25; 6.96), little (156; 6.93), overall (34; 6.90), emotional (21; 6.82), minimal (13; 6.70), marginal (16; 6.69), uneven (11; 6.62)
5	*FRUIT*	citrus (49; 9.81), fresh (182; 9.80), kiwi (27; 8.95), ripe (25; 8.73), passion (23; 8.65), forbidden (19; 8.41), exotic (20; 8.00), tinned (13; 7.84), cereal (14; 7.83), vegetable (17; 7.77), unripe (11; 7.74), canned (11; 7.63), candied (8; 7.27), rotten (10; 7.26), bore (8. 7.23), flower (13; 7.20), fleshy (8; 7.18), soft (25; 7.16), tropical (12; 7.05), vine (7; 7.00), meat (9; 6.88), ugli (6; 6.87), bread (8; 6.85), stewed (6; 6.85), bear (7; 6.81)
6	*OUTCOME*	learning (48; 9.52), likely (74; 9.42), eventual (29; 8.57), satisfactory (29; 8.42), logical (29; 8.21), successful (54; 7.89), inevitable (18; 7.84), favourable (14; 7.56), longterm (9; 7.50), positive (38; 7.44), possible (55; 7.36), final (65; 7.31), ultimate (16; 7.18), chosen (7; 7.07), desirable (8; 6.93), happy (16; 6.85), probable (7; 6.85), policy (24; 6.80), tragic (8; 6.78), unsatisfactory (6; 6.74), behavioural (7; 6.60), disastrous (6; 6.48), clinical (11; 6.47), unexpected (8; 6.45), intended (5; 6.44)
7	*AFTERAMTH*	silage (3; 9.38), immediate (70; 8.71), hay (2; 7.29), sad (2; 5.34), gulf (2; 5.15), election (2; 4.49), bloody (2; 3.84), war (2; 3.62)
8	*SEQUEL*	long-awaited (2; 7.99), logical (3; 5.79), inevitable (2; 5.67), immediate (7; 5.40), interesting (2; 3.85), possible (2; 2.78), own (2; 0.25)
9	*BY-PRODUCT*	undesired (2; 9.24), corrosion (2; 8.81), gaseous (2; 8.46), incidental (3; 8.42), unavoidable (2; 8.02), accidental (3; 7.44), intriguing (2; 7.18), harmful (2; 7.11), inevitable (5; 6.99), unfortunate (3; 6.66), valuable (2; 4.68), product (2; 4.35), gas (2; 4.20), important (9; 3.95), useful (2; 3.86)
10	*UPSHOT*	unsettling (1; 8.25), deleterious (1; 8.24), challenging (1; 6.58), re (1; 6.22), logical (1; 4.27), practical (3; 4.03), moral (1; 2.93), certain (1; 0.94), main (1; 0.52)
11	*END-PRODUCT*	higher-quality (1; 10.14), presentable (1; 9.44), saleable (1; 8.81), insoluble (1; 8.02), predictable (1; 6.27), desirable (1; 5.64), visible (1; 5.03), identical (1; 5.02), acceptable (1; 4.63), stable (1; 4.26), useful (2; 3.89), design (1; 3.36), beautiful (1; 2.71), traditional (1; 1.96), simple (1; 1.70), final (1; 1.46), same (2; 0.45), only (1; 0.45)

Table 5.4 Collocates (as modifiers) of the lemmas

Table 5.4 shows that each word in query has elicited a long list of collocates as modifiers. The collocates listed here are all statistically significant. It can be seen that the words under investigation share some collocates except for *FRUIT*, which is usually used literally, as we can see the collocates such as *citrus, fresh, kiwi, ripe, passion, exotic* and *tropical*. The use of *FRUIT* in the metaphorical sense (i.e. the outcome of a certain happening, event or action) will be discussed later, but for now it may suggest that *FRUIT* is normally not considered as synonymous to the other words in the set.

The data here is distorted by the polysemous use of *FRUIT*. If we eliminate the physical sense of 'the soft part containing seeds that is produced by a plant' with the identification of the collocates, then we are left with *forbidden, bore* and *bear*. We also see that it is necessary to separate the singular and plural forms of the lemma because we could actually eliminate instances of the literal use of *FRUIT* by only looking at the singular and plural forms, and that the metaphorical sense of *FRUIT* does not elicit many collocates. What seems also worth mentioning here is the related issue of hyponymy. In the Table 5.4, we can see that the collocates including *kiwi, cereal, vegetable, vine, meat, ugli* and *bread* are related to hyponyms of *fruit*; therefore we also need to eliminate the instances of hyponyms

use. However even after we eliminate the cases both in polysemy and hyponymy, we find that *fruit* still does not have shared collocates with other words in query.

Collocates	shared by…
adverse	*EFFECT, IMPACT, CONSEQUENCE*
profound	*EFFECT, IMPACT, CONSEQUENCE*
immediate	*EFFECT, IMPACT, SEQUEL, AFTERMATH*
significant	*EFFECT, IMPACT,*
direct	*EFFECT, IMPACT, RESULT, CONSEQUENCE*
dramatic	*EFFECT, IMPACT,*
devastating	*EFFECT, IMPACT,*
overall	*EFFECT, IMPACT, RESULT,*
long-term	*EFFECT, CONSEQUENCE*
damaging	*EFFECT, CONSEQUENCE*
possible	*EFFECT, CONSEQUENCE, OUTCOME, SEQUEL*
harmful	*EFFECT, CONSEQUENCE, BY-PRODUCT*
net	*EFFECT, RESULT,*
negative	*EFFECT, IMPACT, RESULT, CONSEQUENCE,*
positive	*EFFECT, RESULT, OUTCOME*
end	*IMPACT, RESULT,*
election	*RESULT, AFTERMATH*
final	*RESULT, OUTCOME, END-PRODUCT*
inevitable	*RESULT, CONSEQUENCE, OUTCOME, SEQUEL, BY-PRODUCT*
satisfactory	*RESULT, OUTCOME,*
disastrous	*RESULT, CONSEQUENCE, OUTCOME,*
same	*RESULT, END-PRODUCT*
environmental	*IMPACT, CONSEQUENCE*
likely	*IMPACT, CONSEQUENCE, OUTCOME*
unfortunate	*CONSEQUENCE, BY-PRODUCT*
logical	*CONSEQUENCE, OUTCOME, SEQUEL, UPSHOT*
tragic	*CONSEQUENCE, OUTCOME*
important	*CONSEQUENCE, BY-PRODUCT*
practical	*CONSEQUENCE, UPSHOT*
desirable	*OUTCOME, END-PRODUCT*
useful	*BY-PRODUCT, END-PRODUCT*

Table 5.5 Collocates shared by the words in query

All the shared collocates as modifiers are listed in Table 5.5. It can be seen that the shared collocates among the words in an intertwined manner indicate approximation in meaning of interlinked senses. For example, *RESULT, CONSEQUENCE, OUTCOME, SEQUEL* and *BY-PRODUCT* share the collocate *inevitable*; *EFFECT, IMPACT* and *CONSEQUENCE* share *adverse; EFFECT, IMPACT, RESULT* and *CONSEQUENCE* share *direct; EFFECT, IMPACT, AFTERMATH* and *SEQUEL* share *immediate*.

From Table 5.5, we may conclude that the word *EFFECT* seems to have more shared collocates compared with the other words in the set. This in turn seems to suggest that the meaning of *EFFECT* is more general and that the word can be used in more contexts. As mentioned before, the situation is complicate with *FRUIT*, because even if we eliminate the polysemous use, we still do not find many shared collocates with other words in query, which may suggest that *FRUIT* does not share much closeness in meaning/sense with other words in query.

From the shared collocates as modifiers, we could see convergence among the words under investigation as 'you shall judge a word by the company it keeps' (Firth, 1957). For instance, *CONSEQUENCE, OUTCOME, SEQUEL* and *UPSHOT* share the modifier collocate *logical*, which somehow indicates that these four words in query share similarities when it comes to the logical aspect of 'result'. Also *CONSEQUENCE* and *OUTCOME* share *tragic*, a modifier collocate that *SEQUEL* and *UPSHOT* do not share; it suggests that there is a negative association with *CONSEQUENCE* and *OUTCOME*, and none with *SEQUEL* and *UPSHOT*.

5.2.3.2.2. verbs of which the words in query function as Sbject

Next I looked at the verbs where the candidate words function as Object of the clause in question. Table 5.6 lists all the verb collocates with the words in query as Object and Table 5.7 lists the shared collocates.

Rank	Lemma	Collocates (as object of the query words) (Frequency; significance)
1	*RESULT*	achieve (314; 9.16), obtain (277; 9.05), produce (490; 9.04), publish (151; 8.23), report (137; 8.16), yield (88; 8.16), interpret (76; 7.86), desire (67; 7.78), announce (93; 7.70), show (207; 7.65), compare (88; 7.64), analyse (56; 7.35), present (83; 7.30), express (76; 7.20), await (45; 7.15), confirm (49; 7.01), expect (76; 6.91), get (312; 6.87), give (300; 6.83), predict (34; 6.71), summarise (30; 6.70), be (2523; 6.57), affect (49; 6.49), explain (39; 6.44), improve (45; 6.32)
2	*EFFECT*	have (4935; 9.20), produce (307; 8.17), desire (119; 8.08), assess (115; 7.79), achieve (137; 7.65), examine (113; 7.60), study (107; 7.58), exert (77; 7.48), consider (132; 7.32), show (178; 7.26), create (135; 7.26), investigate (74; 7.19), give (398; 7.17), suffer (73; 7.04), take (456; 7.01), evaluate (51; 6.82), offset (43; 6.67), observe (44; 6.51), reduce (75; 6.49), feel (68; 6.42), determine (51; 6.39), measure (43; 6.37), limit (43; 6.32), counteract (31; 6.24), explain (43; 6.21)

3	*CONSEQUENCE*	suffer (44; 7.65), fear (13; 6.92), avoid (31; 6.80), foresee (8; 6.67), predict (11; 6.52), mitigate (6; 6.37), have (601; 6.22), explore (11, 6.18), face (25; 6.17), examine (17; 6.11), consider (33; 6.08), anticipate (6; 5.92), escape (7; 5.83), assess (10; 5.75), risk (5; 5.75), investigate (9; 5.70), accept (19; 5.67), ignore (9; 5.64), evaluate (5; 5.45), understand (12; 5.40), analyse (5; 5.16), define (8; 5.04), experience (6; 4.98), illustrate (5; 4.92), handle (5; 4.91)
4	*IMPACT*	assess (130; 9.23), minimise (35; 8.14), examine (60; 7.75), minimize (22; 7.62), have (1221; 7.23), evaluate (21; 7.17), soften (16; 7.15), make (474; 6.95), lessen (13; 6.89), reduce (63; 6.88), measure (24; 6.79), cushion (11; 6.72), consider (54; 6.69), exert (12; 6.57), predict (12; 6.30), reflect (24; 6.30), appreciate (12; 6.25), investigate (15; 6.19), limit (16; 6.08), offset (8; 6.07), mitigate (7; 6.06), survive (12; 6.02), diminish (8; 6.00), analyse (11; 6.00), feel (31; 5.99)
5	*FRUIT*	dry (75; 9.85), bear (139; 8.96), eat (58; 8.16), taste (10; 7.60), pick (21; 7.59), ripen (6; 7.25), crystalise (6; 7.23), harvest (6; 7.11), peel (6; 7.02), rot (6; 7.00), enjoy (34; 6.96), reap (6; 6.96), pluck (5; 6.78), pile (5; 6.71), soak (5; 6.70), grow (27; 6.61), chop (5; 6.32), forbid (5; 6.27), sell (21; 5.95), wash (6; 5.87), store (5; 5.78), fall (6; 5.77), buy (22; 5.65), produce (29; 5.38), collect (7; 5.23)
6	*OUTCOME*	predict (44; 8.81), await (38; 8.71), pend (27; 8.68), influence (53; 8.45), desire (19; 7.87), determine (58; 7.85), affect (60; 7.56), learn (41; 7.28), evaluate (11; 6.89), decide (10; 6.52), prejudge 94; 6.44), assess (14; 6.42), achieve (26; 6.33), prejudice (4; 6.26), record (15; 6.19), regret (4; 6.05), forecast (4; 6.03), intend (8; 6.00), secure (10; 5.93), monitor (6; 5.75), produce (37; 5.71), anticipate (4; 5.68), imagine (6; 5.66), expect (20; 5.62), improve (16. 5.56)
7	*AFTERMATH*	survey (3; 6.79), discuss (2; 2.99), leave (2; 1.04), follow (2; 1.01), see (3; 0.26)
8	*SEQUEL*	commission (2; 0.31), write (8; 4.19), plan (2; 3.57), describe (2; 2.58), produce (2; 1.59), do (5; 1.24), know (2; 1.00), make (6; 0.68), be (33; 0.32)
9	*BY-PRODUCT*	produce (4; 2.60), form (2; 2.39), be (66; 1.32), become (2; 1.29)
10	*UPSHOT*	formulate (2; 6.00), confine (1; 4.47), predict (1; 4.23), interpret (1; 4.14), treat (1; 2.63), obtain (1; 1.99)
11	*END-PRODUCT*	synthesize (1; 7.81), desire (2; 5.78), argue (1; 4.69), interpret (1; 4.14), handle (1; 3.32), obtain (1; 1.99), represent (1; 1.63), leave (1; 0.04)

Table 5.6 Verbs collocates with the words in query as Object

The verb collocates show some features of meanings/senses of the words in query as Objects. For example, by looking at the verb collocates, we know that we could '*achieve, obtain, produce, publish, report, yield, interpret, desire, announce, show, compare, analyse, present, express, await, confirm, expect, get, give, predict, summarise, affect, explain and improve* a *result*'; and also we could *'have, produce, desire, assess, achieve, examine, study, exert, consider, show, create, investigate, give, suffer, take, evaluate, offset, observe, reduce, feel, determine, measure, limit, counteract, and explain* an *effect* (see Table 5.6). Comparing the verb collocates of *RESULT* and *EFFECT*, we find that the two words under investigation share the similarity that they both collocate with verbs *observe, produce, desire* and *show* though with different frequencies and significance. The overlapped verb collocates are also found with other words in query, which seems to show convergence among the candidate words in terms of their association with verbs when functioning as Objects.

Collocates (verb)	Shared by…
have	*EFFECT, IMPACT, CONSEQUENCE*
produce	*EFFECT, RESULT, FRUIT, OUTCOME, SEQUEL, BY-PRODUCT*
desire	*EFFECT, RESULT, OUTCOME, END-PRODUCT*
assess	*EFFECT, IMPACT, CONSEQUENCE, OUTCOME*
achieve	*EFFECT, RESULT, OUTCOME*
examine	*EFFECT, IMPACT, CONSEQUENCE*
exert	*EFFECT, IMPACT*
consider	*EFFECT, IMPACT, CONSEQUENCE,*
show	*EFFECT, RESULT*
investigate	*EFFECT, IMPACT, CONSEQUENCE,*
give	*EFFECT, RESULT*
suffer	*EFFECT, CONSEQUENCE,*
evaluate	*EFFECT, IMPACT, CONSEQUENCE, OUTCOME*
offset	*EFFECT, IMPACT,*
reduce	*EFFECT, IMPACT,*
feel	*EFFECT, IMPACT,*
determine	*EFFECT, OUTCOME*
measure	*EFFECT, IMPACT,*
limit	*EFFECT, IMPACT,*
explain	*EFFECT, RESULT*
obtain	*RESULT, UPSHOT, END-PRODUCT*
interpret	*RESULT, UPSHOT, END-PRODUCT*
analyse	*RESULT, IMPACT, CONSEQUENCE*
present	*RESULT, END-PRODUCT*
await	*RESULT, OUTCOME*
expect	*RESULT, OUTCOME*
predict	*RESULT, CONSEQUENCE, OUTCOME, IMPACT, UPSHOT*
be	*RESULT, SEQUEL, BY-PRODUCT*
affect	*RESULT, OUTCOME*
improve	*RESULT, OUTCOME*
make	*IMPACT, SEQUEL*
mitigate	*IMPACT, CONSEQUENCE*
anticipate	*CONSEQUENCE, OUTCOME*
handle	*CONSEQUENCE, END-PRODUCT*
leave	*AFTERMATH, END-PRODUCT*

Table 5.7 Shared verb collocates with the words in query as Object

Two points are worth mentioning. First, candidate synonyms seem to attract other candidate synonyms. For example, *assess* and *evaluate*, which both appear as collocating verbs, seem to share similar meanings. Interestingly, Jones (2002) also finds that antonyms attract other antonyms. When discussing 'ancillary antonymy', Jones (2002) labels *quickly* and *slowly* as 'A-pair' and *now* and *later* 'B-pair' in sentence 1 below. He then discovered a number of different types of B-pair such as antonyms (sentence 2), synonyms (sentence 3) and metonyms (sentence 4).

4. If so, unemployment may rise more **quickly** *now*, but more **slowly** *later*.

5. He also suggests discipline should be tailored differently, saying **extroverts** are most motivated by *reward* while **introverts** respond more to *punishment*.

6. Then, and now, the Royal Festival Hall is a cool, rather clinical building that it is **easy** to *respect* and **difficult** to *love*.

7. But a couple of Libyans are only likely to be **small** *minnows* in a very **large** *pond*.

Secondly, it will be noted that two of the verb collocates of our candidate synonyms -- *mitigate, minimize* and *minimise* (American and UK spellings differences) -- can be grouped into a semantic set. The categorization of semantic sets will be discussion is section 5.2.3.3.

5.2.3.2.3. verbs collocating with the words in query where the latter function as Subject in the clause in question

After looking at the verbs where the words in query function as Object, I turn to the verbs where the words in query function as Subject. Table 5.8 shows all the verb collocates with the words in query as Subject. The result shows that *EFFECT* and *RESULT* have more verb collocates with the words in query as Subject than do the other words in the set. In other words when the words *EFFECT* and *RESULT* are the Subject of their clauses, a variety of verbs are capable of functioning as Predicator. On the other hand, *AFTERMATH* and *BY-PRODUCT* do not have any verb collocates when they function as Subject of clause, which might indicate that the two words are rarely used as Subject of a clause. Again, the features of the sense of the words under investigation can be seen by their verb collocates (i.e. the company they keep). For instance, a '*result*' can '*indicate, show, suggest, confirm, demonstrate, support, obtain, be, reflect, encourage, present, follow, reveal, prove, mean, give, agree, provide, depend, achieve, imply, report , compare, improve* and *illustrate*' (see Table 5.8). And an '*effect*' can '*occur, depend, outweigh, observe, cause, happen, last, mean, arise, vary, influence, reduce, rsult, operate, require, become, apply, increase, produce, include, create, do* and *seem* (also

see Table 5.8). The overlapping verb collocates (*depend*, *mean* and *be*) of *RESULT* and *EFFECT* show that the two words share similarities in their predicates when functioning as subjects.

Rank	Lemma	Verbs with the words in query as subject
1	*RESULT*	indicate (126; 8.49), show (285, 8.48), suggest (179; 8.43), confirm (69; 7.83), demonstrate (40; 7.24), support (43; 6.90), obtain (26; 6.69), be (4020; 6.57), reflect (26; 6.44), encourage (21; 6.30), present (22; 6.13), follow (48; 5.94), reveal (17; 5.71), prove (16; 5.63), mean (22; 5.62), give (43; 5.62), agree (20; 5.54), provide (29; 5.52), depend (15; 5.51), achieve (12; 5.42), imply (11; 5.42), report (18, 5.41), compare (9; 5.34), improve (10; 5.30), illustrate (10; 5.30)
2	*EFFECT*	occur (44; 7.18), depend (19; 6.36), outweigh (9; 6.33), observe (11; 6.20), cause (20; 5.97), happen (14; 5.91), last (8; 5.87), mean (20; 5.85), arise (15; 5.81), vary (10; 5.8), influence (9; 5.75), reduce (9; 5.73), RESULT (8; 5.59), be (1946; 5.53), operate (10; 5.45), require (13; 5.37), wear (8; 5.36), become (32; 5.23), apply (8; 5.22), increase (9; 5.14), produce (11; 4.95), include (19; 4.95), create (7; 4.90), do (72; 4.82), seem (18; 4.74)
3	*CONSEQUENCE*	arise (7; 5.27), occur (8; 5.20), follow (20; 5.07), be (337; 3.00), have (55; 2.08), do (5; 1.02)
4	*IMPACT*	occur (6; 4.80), cause (6; 4.71), do (19; 2.95), be (231; 2.45), come (6; 2.09), have (50; 1.94)
5	*FRUIT*	ripen (5; 8.71), grow (11; 5.62), fall (8; 4.57), be (158; 1.90), have (27; 1.05)
6	*OUTCOME*	depend (11; 6.22), reflect (6; 5.68), seem (9; 3.94), be 9481; 3.51), have (47; 1,85), go (4; 1.55), do (5; 1.02)
7	*AFTERMATH*	/
8	*SEQUEL*	suffer (2; 3.90), come (2; 0.52)
9	*BY-PRODUCT*	/
10	*UPSHOT*	seem (1; 0.81), be (60; 0.51)
11	*END-PRODUCT*	arrive (1; 2.73)

Table 5.8 Verb collocates with the words in query as Subject

Table 5.9 lists all the shared verb collocates with the words in query as Subject. The word *EFFECT* appears most frequently with other words sharing the verb collocates, which again seems to indicate that the meaning of *EFFECT* is more general and can be used in more contexts. As before *FRUIT* does not seem to share verb collocates except *be* and *have* with other words in query, which seems to again suggest that *FRUIT* do not share closeness of meanings or senses with other words in query.

The shared collocates seem to suggest the degrees of similarity and difference among these candidate words. Quirk (1967) coins the term ‘cline’ for a situation where the criterion is not needed to sort words into two classes. We saw that in Chapter 1 the discussion of distinction of verb and noun in Chinese and also in English we cannot make an absolute decision what position they are along the line. The shared collocates of the words here suggest we have a cline of synonymy.

Collocates (verb)	Shared by…
occur	*EFFECT, IMPACT, CONSEQUENCE*
depend	*EFFECT, RESULT, OUTCOME*
cause	*EFFECT, IMPACT*
mean	*EFFECT, RESULT*
arise	*EFFECT, CONSEQUENCE*
be	*EFFECT, RESULT, IMPACT, CONSEQUENCE, FRUIT, OUTCOME, UPSHOT*
do	*EFFECT, IMPACT, CONSEQUENCE, OUTCOME*
seem	*EFFECT, OUTCOME, UPSHOT*
follow	*RESULT, CONSEQUENCE*
come	*IMPACT, SEQUEL*
have	*IMPACT, CONSEQUENCE, FRUIT, OUTCOME*
reflect	*RESULT, OUTCOME*

Table 5.9 Shared verb collocates with the words in query as Subject

5.2.3.2.4. words which appear in the structure of 'the query words + *and/or* + noun'

Next I looked at the nouns and also the head nouns in the noun phrases that appear in the structure 'the query words + *and/or* + noun' with the words in query. I also eliminated the cases when noun is in a postmodifier position. Look at the following examples:

Example 5.24:

At the time these recordings are made the infant cannot yet relate *cause and* ***effect***, so approval and reward for good behaviour tend to produce confusion.

Example 5.25:

Before having the HIV antibody test, a person should consider the implications of telling other people about the *test and* its ***result***.

Example 5.26:

Implementation will require all courses to be specified in the new unit-based format, with candidates' performance detailed in terms of ***outcomes*** *and* performance *criteria.*

Example 5.27:

Looking at your diary again, can you see any kind of pattern in the ***consequences or outcomes*** to these trying confrontations?

In example 5.24, *cause* is in a parallel structure with *effect*, linked by the word *and*. And similar situations may be found for *test* with *result* (example 5.25), *outcome*s with *criteria* (example 5.26) and *consequences* with *outcomes* (example 5.27).

Rank	Lemma	Nouns in the structure of 'words in query + and/or + Noun'
1	*RESULT*	test (23, 7.37), performance (15; 6.64), conclusion (10; 6.52), number (23; 6.46), report (17; 6.46), action (16; 6.45), cause (11; 6.44), study (15; 6.27), method (11; 6.05), background (8; 6.03), score (7; 6.02), datum (9; 6.00), experiment (7; 5.95), theory (10; 5.92), steel (7; 5.89), asset (7; 5.89), dividend (6; 5.89), cent (7; 5.86), finding (6; 5.81), process (10; 5.78), detail (8; 5.76), technique (8; 5.75), rate (11; 5.72), year (22; 5.68), effort (7; 5.66)
2	*EFFECT*	cause (186; 10.52), circumstance (36; 8.07), meaning (26; 7.68), nature (26; 7.21), purpose (19; 7.00), implication (11; 6.60), force (17; 6.57), interaction (8; 6.22), premise (8; 6.19), change (16; 6.18), music (13; 6.16), impact (8; 6.16), sound (10; 6.14), factor (9; 6.08), appearance (8; 6.04), scale (8; 6.01), use (12; 5.92), war (11; 5.87), extent (7; 5.80), colour (10; 5.78), lighting (6; 5.78), statement (7; 5.75), intention (6; 5.75), cost (12; 5.74), pattern (8; 5.70)
3	*CONSEQUENCE*	antecedent (22; 9.56), cause (40; 9.34), determinant (6; 7.69), implication (9; 7.69), outcome (5; 6.86), behavior (8; 6.41), illness (5; 6.38), nature (8; 6.18), term (5; 5.54), event (5; 5.46), action (5; 5.40), condition (6; 5.32), cost (6; 5.23), course (8; 5.23), use (5; 5.23), level (5; 5.04), change (5; 5.00), problem (5; 4.89), people (6; 4.16)
4	*IMPACT*	immediacy (5; 7.89), effectiveness (8; 7.89), scale (10; 7.61), significance (5; 7.28), effect (8; 6.16), policy (10; 5.89), change (8; 5.75), development (6; 4.78), way (5; 4.74)
5	*FRUIT*	vegetable (415; 12.17), flower (131; 10.21), nut (56; 9.59), veg (38; 9.38), cereal (20; 8.31), seed (21; 8.16), salad (18; 8.00), berry (15; 7.89), leave (20; 7.84), meat (19; 7.76), bread (22; 7.75), cheese (19; 7.70), yogurt (12; 7.70), apple (12; 7.35), tomato (10; 7.12), yoghurt (8; 7.09), wine (14; 7.00), juice (9; 6.98), cake 910; 6.97), milk (10; 6.86), orange (8; 6.84), fish (14; 6.82), cream (9; 6.78), market (16; 6.69), foliage (6; 6.65)
6	*OUTCOME*	criterion (12; 8.68), process (23; 7.79), variable (4; 7.23), effectiveness (4; 7.07), objective (6; 6.88), CONSEQUENCE (5; 6;86), procedure (6; 6.30), decision (5; 6.14), content (4; 5.94), project (4; 5.75), care (5; 5.49), quality (5; 5.46), use (5; 5.37), cost (4; 4.77), course (5; 4.63), government (5; 4.42)
7	*AFTERMATH*	hay (3; 8.15), invasion (2; 7.61), preparation (2; 6.37), war (8; 6.35), case (4; 4.73), event (2; 4.51)
8	*SEQUEL*	Prequel (2; 10.44), Zenda (2; 10.32), Rupert (2; 9.20), movie (2; 7.56)
9	*BY-PRODUCT*	air (2; 5.20), product (3; 4.86)
10	*UPSHOT*	gist (2; 11.47), standardization (1; 9.91), mace (1; 8.55), heath (1; 6.31), nothing (2; 5.12), summer (1; 5.01), purpose (1; 4.15), plan (1; 3.89), paper (1; 3.46), RESULT (1; 3.43)
11	*END-PRODUCT*	block (1; 5.61), exercise (1; 5.10), goal (1; 4.91), responsibility (1; 4.31), action (1; 3.46), example (1; 1.95)

Table 5.10 Words which appear in the structure of 'the words in query + *and/or* + Noun'

Table 5.10 lists all the nouns in this structure with the words in query. Again with each noun two figures are provided, one being the frequency of the word and the other being the significance of the collocational association. From the table, it can be seen that the top significant associations are *FRUIT* and *vegetable* (415; 12.17), and *EFFECT* and *cause* (186; 10.52). It has been mentioned in Chapter 4 that into the psychological reality of synonymy, to the prompt word the participants provided a variety

of words, some of which may not be considered as synonyms. Two of these pairs of noun collocates (*fruit* and *vegetable*, and *effect* and *cause*) may offer some explanations of the result we obtained in the experiment. The corpus analysis seems to suggest that these two pairs are significantly collocated, and therefore people have been primed to their associations and offered one as the candidate synonym to the other within the limited time in the experiment, even though if given more time, the participants would realise the words are not synonymous.

Again, we may pick up on the confusing effects of polysemy. The issue is that a polysemous word can by definition only be a potential synonym in one of the senses. After all, a polysemous word (again by definition) is not even a synonym of itself.

However some of the words in the table may arouse doubt as there seems to be only one or two instances. Note that the words are provided based on significance of collocational association, scored in logDice. In spite of the small number of occurrences, the collocational association between these nouns and the query words in this structure are still statistically significant, and therefore worth our attention. For example, for *SEQUEL*, the words *movie* and *Zenda* are in the list of noun collocation in the structure. Although small in number, these occurrences seem to suggest that *SEQUEL* has the sense of 'a book, film, or play that continues the story of a previous book', which is not shared with other query words.

Example 5.28:

Pesci, a bungling burglar in the *movie* and its upcoming ***sequel***, gave up singing for acting 15 years ago.

Example 5.29:

The Prisoner of *Zenda* and its ***sequel*** certainly bear witness to their author's craftsmanship.

Table 5.11 lists all the shared noun collocates in the structure of 'the query word + *and/or* + noun'. It is worth noting that the criterion that the nouns are shared eliminates the words which may suggest unrelated sense (for example, *Zenda* for *SEQUEL*) or apparent rubbish (*mace* for *UPSHOT*). Again the shared noun collocates suggest the closeness of the words in query. For example, *EFFECT*, *RESULT* and *CONSEQUENCE* share the noun collocate *cause*; while *EFFECT* and *CONSEQUENCE* share the noun collocates *nature*, *implication* and *cost*. In other words, there is more similarities between *EFFECT* and *CONSEQUENCE* than between either of these words and *RESULT* when they are used in the structure of 'the query word + *and/or* + noun'.

Collocates	shared by…
cause	*EFFECT, RESULT, CONSEQUENCE*
nature	*EFFECT, CONSEQUENCE*
purpose	*EFFECT, UPSHOT*
implication	*EFFECT, CONSEQUENCE*
change	*EFFECT, IMPACT, CONSEQUENCE*
scale	*EFFECT, IMPACT*
use	*EFFECT, CONSEQUENCE, OUTCOME*
war	*EFFECT, AFTERMATH*
cost	*EFFECT, CONSEQUENCE*
action	*RESULT, CONSEQUENCE, END-PRODUCT*
process	*RESULT, OUTCOME*
effectiveness	*EFFECT, OUTCOME*
event	*CONSEQUENCE, AFTERMATH*
course	*CONSEQUENCE, OUTCOME*

Table 5.11 Shared nouns which appear in the structure of 'the words in query + and/or + noun'

5.2.3.2.5. prepositions which occur within L3 and R3 of the words in query

Next the prepositions with which the query words collocate are looked at. Table 5.12 lists all prepositions which occur within 3-word-span on the left and right sides of the words in query. Note again both figures of frequency and significance are provided in the table. Due to the high frequency of prepositions in the corpus, the significance of propositions collocating with the query words is relatively low; therefore I only concentrate on the prepositions whose significance is above 0.03 (in bold in Table 5.12).

It will be apparent from Table 5.12 that the left co-occurrences of the words under investigation with the top significance are *in* with *AFTERMATH* (250; 0.36), *as* with *BY-PRODUCT* (39; 0.15), *as* with *RESULT* (3928; 0.12) and *of* with *FRUIT* (527; 0.11). Our introspection would justify the phrase 'in the aftermath', 'as a result' and 'fruit of …'; the low appearance of *BY-PRODUCT* in the corpus provides a possible explanation as to why we are not primed for this co-occurrence in the way that we are for 'as a result'. On the other hand, all the words in query have right co-occurrences of *of* with a collocational significance over 0.03 except *SEQUEL* (only 0.01), whose top right co-occurrence is *to* (0.26). It seems also to suggest that *SEQUEL* shares less similarity with other candidate words in terms of its right co-occurrences of prepositions.

Rank	Lemma	Left Co-occurrences	Right Co-occurrences
1	*RESULT*	**as (3928; 0.12), of (1209; 0.04),** with (621; 0.02), to (294; 0.01), on (249; 0.01), for (248; 0.01), in (230; 0.01), by (202, 0.01), from (103; 0.00), about (88; 0.00), at (81; 0.00), if (49; 0.00), than (39; 0.00), between (39; 0.00)	**of (9339; 0.28), in (1327; 0.04),** from (642; 0.02), for (363; 0.01), with (130; 0.00), on (117; 0.00), at (90; 0.00), to (81; 0.00), as (62; 0.00), by (40; 0.00), than (35; 0.00)
2	*EFFECT*	**of (1605; 0.05), to (1599; 0.05),** in (783, 0.02), into (544; 0.02), with (527; 0.02), for (437; 0.01), about (328; 0.01), on (309; 0.01), from (290; 0.01), by (196; 0.01), as (119; 0.00), at (93; 0.00), without (61; 0.00), than (44; 0.00)	**of (7990; 0.24), on (3713; 0.11),** in (577; 0.02), from (343; 0.01), upon (288; 0.01), to (254; 0.01), for (120; 0.00), as (109; 0.00), at (87; 0.00), with (72; 0.00), by (63; 0.00)
3	*CONSEQUENCE*	**of (519; 0.07), as (388; 0.05),** in (174; 0.02), with (162; 0.02), about (97; 0.01), to (83; 0.01), for (81; 0.01), from (52; 0.01), on (51; 0.01), by (30; 0.00), at (23; 0.00), without (20; 0.00), through (14; 0.00), against (13; 0.00), over (10; 0.00), into (9; 0.00)	**of (2656; 0.34), for (407; 0.05),** in (98; 0.01), to (38; 0.00), on (21; 0.00), if (14; 0.00), as (12; 0.00), from (10; 0.00), at (8; 0.00)
4	*IMPACT*	**of (525; 0.07),** about (128; 0.02), to (126; 0.02), on (125; 0.02), with (83; 0.01), by (64; 0.01), for (59; 0.01), at (59; 0.01), in (58; 0.01), from (55; 0.01), under (41; 0.01), through (22; 0.00), into (21; 0.00), as (18; 0.00), over(16; 0.00), before (13; 0.00),	**of (2298; 0.30),** on (1522; 0.20), upon (145; 0.02), in (128; 0.02), with (30; 0.00), at (25; 0.00), to (15; 0.00), than (15; 0.00), for (13; 0.00)
5	*FRUIT*	**of (527; 0.11),** with (98; 0.02), in (66; 0.01), for (51; 0.01), to (39; 0.01), on (39; 0.01), as (37; 0.01), from (32; 0.01), like (27; 0.01), by (16; 0.00), over (12; 0.00), into (11; 0.00), at (5; 0.00)	**of (397; 0.08),** in (94; 0.02), from (35; 0.01), on (32; 0.01), with (32; 0.01), for (31; 0.01), to (16; 0.00), at (11; 0.00), like (8; 0.00), into (7; 0.00), as (7; 0.00)
6	*OUTCOME*	**of (313; 0.07), on (145; 0.03), to (130; 0.03),** about (97; 0.02), as (74; 0.02), for (61; 0.01), at (30; 0.01), by (28; 0.01), towards (12; 0.00), upon (12; 0.00), from (8; 0.00), over (8; 0.00), if (8; 0.00), until (8; 0.00)	**of (1347; 0.29),** in (100; 0.02), for (79; 0.02), to (34; 0.01), with (25; 0.01), from (19; 0.00), as (14; 0.00), on (13; 0.00), at (11; 0.00)
7	*AFTERMATH*	**in (250; 0.36)**, with (14; 0.02), of (11; 0.02), to (8; 0.01), on (3; 0.00), from (3; 0.00), by (3; 0.00), about (3; 0.00), at (2; 0.00)	**of (457; 0.66)**, in (2; 0.00), to (2; 0.00)
8	*SEQUEL*	**in (16; 0.05), of (12; 0.04),** for (7; 0.02), as (5; 0.02), to (4; 0.01), from (3; 0.01), with (3; 0.01)	**to (79; 0.26)**, in (4; 0.01), of (2; 0.01)
9	*BY-PRODUCT*	**as (39; 0.15), of (9; 0.04),** to (2; 0.01)	**of (134; 0.53),** from (5; 0.02), in (2; 0.01), on (2; 0.01)
10	*UPSHOT*	of (1; 0.01), to (1; 0.01), at (1; 0.01), if (1; 0.01)	**of (29; 0.19),** for (1; 0.01)
11	*END-PRODUCT*	**of (3; 0.05), at (3; 0.05), in (2; 0.03), for (2; 0.03), with (2; 0.03),** to (1; 0.02), from (1; 0.02), into (1; 0.02), as (1; 0.02), through (1; 0.02)	**of (18; 0.31)**

Table 5.12 Prepositions which occur on the left and right of the words in query

Table 5.13 lists shared prepositions (significance above 0.03) on the left and right of the words in query, in which we can see some patterns. For instance, the preposition *as* occurs with significance on the left side of *RESULT*, *CONSEQUENCE* and *BY-PRODUCT*, which matches our intuitions since we are familiar with the phrases *as a result* and *as a consequence* (commonly considered as fixed phrases), and maybe also *as* a *BY-PRODUCT* (a co-occurrence overlooked somehow), but not *'as an effect'**. Also the preposition *in* appears on the left side of *AFTERMATH*, *SEQUEL* and *END-PRODUCT*, but not of *EFFECT* or *OUTCOME*.

left co-occurrence	as	*RESULT, CONSEQUENCE, BY-PRODUCT*
	of	*RESULT, EFFECT, CONSEQUENCE, IMPACT, FRUIT, OUTCOME, SEQUEL, END-PRODUCT*
	to	*EFFECT, OUTCOME*
	in	*AFTERMATH, SEQUEL, END-PRODUCT*
right co-occurrence	of	*RESULT, EFFECT, CONSEQUENCE, IMPACT, FRUIT, OUTCOME, AFTERMATH, BY-PRODUCT, UPSHOT, END-PRODUCT*

Table 5.13 Shared prepositions (significance above 0.03) on the left and right of the Words in Query

5.2.3.2.6. noun heads occurring in the structure of 'the query word + *of* + head noun'

In addition I looked at the structure 'words in query (e.g. *EFFECT, OUTCOME…*) + *of* + head noun*'* and with the focus being on the head noun after *of.* Table 5.14 lists all the noun heads occurring in the structure.

Table 5.14 shows that various nouns occur in the structure, from which we could see some features of the words in query. For example, when we talk about result, we usually say the '*result*' of *a 'survey, experiment, test, accident, investigation, study, election, research, analysis, ballot, pressure, injury, inquiry, merger, examination, change, trial, failure, action, decision, testing, exercise, assessment, lack or crisis'*. And '*effect*' is usually of '*recession, change, alcohol, drug, smoking, mutation, calcium, taxation, stress, inflation, radiation, variable, exposure, arousal, toxin, treatment, pollution, war, uncertainty, vitamin, gastrin, warming, unemployment, tax* or *inhibitor*'. Later on I will discuss semantic association, but here we may have noticed that *survey, experiment, test, investigation, research, analysis inquiry, examination, trial, testing* and *assessment* are all to do with research; while *recession, alcohol, drug, smoking, mutation inflation, radiation, exposure, toxin and pollution* are to do with effect of being unhealthy. Note how the collocates of two words are different and but there are some overlaps, for example, collocates which have the meaning that people have a negative attitude including *pressure, stress, failure, lack, crisis, inflation, war* and *uncertainty.*

Although most of these noun heads only occur once or twice, it tells us what types of nouns may appear in the structure and suggest the closeness of meaning/sense among the words in query. The noun collocate *research* is shared by *RESULT* and *FRUIT,* which does indicate the closeness of senses between the two words. The word *SEQUEL* does not have any noun in this structure, which indicates that *SEQUEL* does not usually appear in the structure of '*SEQUEL* + of + Noun'.

Rank	Lemma	Noun heads in the structure of 'the query word + *of* + Noun'
1	*RESULT*	survey, experiment, test, accident, investigation, study, election, research, analysis, ballot, pressure, injury, inquiry, merger, examination, change, trial, failure, action, decision, testing, exercise, assessment, lack, crisis
2	*EFFECT*	recession, change, alcohol, drug, smoking, mutation, calcium, taxation, stress, inflation, radiation, variable, exposure, arousal, toxin, treatment, pollution, war, uncertainty, vitamin, gastrin, warming, unemployment, tax, inhibitor
3	*CONSEQUENCE*	literacy, nullity, breach, failure, instability, action, addiction, neglect, negligence, change, refusal, accident, intercourse, defect, ignorance, decline, drinking, decision, shift, divorce, interaction, inability, inequality, revolution, sin
4	*IMPACT*	recession, warming, deregulation, tunnel, technology, fundholding, change, unemployment, crisis, newcomer, constraint, reform, mining, war, divorce, agriculture, media, redundancy, tax, sanction, merger, policy, revolution, inflation
5	*FRUIT*	endeavour, labour, collaboration, earth, victory, spirit, tree, success, species, effort, research, knowledge, experience, work, action, industry, year
6	*OUTCOME*	deliberation, hypoxaemium, election, inquiry, referendum, negotiation, contest, struggle, summit, pregnancy, experiment, dispute, motivation, proceedings, process, discussion, review, talk, appeal, investigation, battle, vote, enquiry, audit, trial
7	*AFTERMATH*	massacre, emancipation, debacle, plague, uprising, courtship, rioting, disaster, riot, atrocity, war, coup, turmoil, hurricane, explosion, earthquake, revolution, tragedy, Eruption, defeat, signing, liberation, blast
8	*SEQUEL*	/
9	*BY-PRODUCT*	corrosion, tin, electricity, intelligence, revolution, teaching, investigation, process, war, reaction, method, practice, form, industry, activity, study, history, forces, transformation, marriage, postmodernist, photosynthesis, regimes, system, worship, attempt, art, phenomenon, emergency, contact, way, endeavour, work, fission, illness, submission, competition,
10	*UPSHOT*	shenanigan, foray, litigation, calculation, inquiry, negotiation, conflict, session, pressure, trial, budget, matter, discussion, strategy, difficulty, visit, principle, theory, work, view, meeting, report, thing, position, research, action
11	*END-PRODUCT*	glycolysis, collision, pathway, orientation, farming, explosion, exercise, trend, generation, race, procedure, activity, stage, period, work

Table 5.14 Noun heads occurring in the structure of 'words in query (e.g. *EFFECT*, *OUTCOME*…) +*of* +noun'

A number of nouns were shared across the query words in this structure. Table 5.15 lists all the shared noun heads in the *of*-structure. As can be seen from Table 5.15, *EFFECT*, *IMPACT*, *AFTERMATH* and *BY-PRODUCT* share *war*; and *EFFECT* and *IMPACT* share both *war* and *inflation*. It seems that *EFFECT* and *RESULT* have a greater number of shared noun heads. Interestingly, the word *FRUIT* shares the noun head *research* with *RESULT,* and *industry* with *BY-PRODUCT* respectively. This test overcomes the problem of polysemy; therefore to study the metaphorical use of *FRUIT* a further study of *FRUIT* with collocates *research* and *industry* is recommended.

Noun heads	shared by …
recession	*EFFECT, IMPACT*
change	*EFFECT, IMPACT, RESULT, CONSEQUENCE*
inflation	*EFFECT, IMPACT*
war	*EFFECT, IMPACT, AFTERMATH, BY-PRODUCT*
trial	*RESULT, OUTCOME, UPSHOT*
research	*RESULT, FRUIT*
industry	*FRUIT, BY-PRODUCT*
failure	*RESULT, CONSEQUENCE*
decision	*RESULT, CONSEQUENCE*
experiment	*RESULT, OUTCOME*
accident	*RESULT, CONSEQUENCE*
investigation	*RESULT, OUTCOME, BY-PRODUCT*
study	*RESULT, BY-PRODUCT*
election	*RESULT, OUTCOME*
action	*RESULT, CONSEQUENCE, FRUIT*
unemployment	*EFFECT, IMPACT*
tax	*EFFECT, IMPACT*
pressure	*EFFECT, UPSHOT*
inquiry	*RESULT, UPSHOT*
merger	*RESULT, IMPACT*
warming	*EFFECT, IMPACT*
exercise	*RESULT, END-PRODUCT*
crisis	*RESULT, IMPACT*
revolution	*IMPACT, AFTERMATH, BY-PRODUCT, CONSEQUENCE*
defect	*CONSEQUENCE, AFTERMATH*
negotiation	*OUTCOME, UPSHOT*
explosion	*AFTERMATH, END-PRODUCT*

Table 5.15 Shared noun heads in the structure of 'words in query + *of* + noun'

To sum up, by looking at the collocations of these candidate synonyms with different functions, we may come up a preliminary conclusion that the words under investigation share some convergence in one way or another in terms of their collocations, and therefore it is safe to say some words are more synonymous than others in certain contexts, co-texts or surroundings, but it is complicated to decide whether candidate words share enough to justify us in considering them to be synonyms or not. This

finding seems to be consistent with what we have found in the psycholinguistic experiment. Various candidate words were offered as synonyms by the participants in the test; however we could only see the closeness or associations between the prompt words and the candidate words provided by the participants. We can not make absolute decisions whether the prompt words and the candidate words are synonyms or not.

5.2.3.3. Semantic association

The collocational analysis has shown that the words in query share some collocates and it may suggest closeness or similarity of meanings among these words. However sharing collocates is not sufficient to enable us to categorise the words as synonyms and therefore the next step of analysis involves semantic association. Let us begin by considering the most frequent word *EFFECT* (297.0 per million) as an example. The modifier collocates and co-occurrences might be categorised into the following five semantic sets. The first is the LOGIC association:

> The **immediate** effect of the latest decision will be to allow sales of timber already cut.
>
> This will have a **knock-on** effect throughout the economy, and will drive up interest rates generally.
>
> The **long-term** psychological effects of this kind of violence can be devastating.

The second set concerns NEGATIVE association and examples are:

> Probably wondering if it would have some sort of **adverse** effect on his investment!
>
> For the first time Gould came up against the **devastating** effects of unlimited commercial exploitation.
>
> The rigid application of `zoning' policies (where indeed it continues) can have a very **damaging** effect.

The third set comprise items indicating SERIOUSNESS.

> However, the individual circumstances of particular plaintiffs clearly have a **significant** effect upon the assessment of damages.
>
> This has a **dramatic** effect on the information management strategy of the organisation.

The fourth concerns SAME/DIFFERENCE, which can also be categorised as COMPARISON AND CONTRAST.

> Smaller quantities of carboxymethyl cellulose, on the other hand, have just the **opposite** effect, helping to stick montmorillonite particles together.

The fifth is the POSITIVE association, which can be also co-grouped with NEGATIVE into the more general association of EVALUATION.

> A controlled trial comparing 12 to 24 weeks of treatment failed to show any **beneficial** effect of the prolonged therapy.
>
> A recent NASA research document details the **positive** effect that plants have in cleaning the air.

The above categorisation makes use of Hoey's (2005) analysis of semantic association of *RESULT* and *CONSEQUENCE* with the purpose of further comparison of the results.
A similar process of analysis was conducted with each candidate word in the group. A random sample of 300 instances of each lemma was retrieved from the BNC (except for *SEQUEL*, *BY-PRODUCT*, *UPSHOT* and *END-PRODUCT* where the full sample was examined since the hits of these four words were 301, 254, 154 and 57 respectively). The results are summarised in Tables 5.15 and 5.16. Table 5.16 lists the modifier collocates in different semantic sets and Table 5.17 summarises the distribution and percentage of the different semantic sets among the words under investigation.

Semantic Sets	Logic	Evaluation		unexpectedness	seriousness	Comparison/contrast (Same/difference)
		Negative	Positive			
RESULT	direct, final, end, desired, interim, preliminary, inevitable,	negative, poor, disappointing, disastrous,	positive, excellent, satisfactory, good,			better, similar, best,
EFFECT	cumulative, direct, immediate, desired, knock-on, indirect, long-term,	adverse, devastating, damaging, negative, detrimental, harmful, side, deleterious	beneficial, positive,	/	profound, significance, dramatic,	opposite,
CONSEQUENCE	inevitable, logical, direct, long-term	disastrous, dire, unintended, adverse, damaging, tragic, unfortunate, harmful, negative, undesirable, fatal, inflationary, catastrophic, devastating		Unforeseen	far-reaching, serious, profound	better, different
IMPACT	immediate, direct, medium-term, major, long-term, likely, potential,	undesirable, adverse,	positive,		remarkable, great, serious, significant, dramatic, tremendous, profound, considerable	greatest, greater, same, maximum, extra, differential
FRUIT	/	/	/	/	/	/
OUTCOME	eventual, logical, desired, long-term, intended, inevitable, final, ultimate, likeliest, expected, predicted, probable, determinate	tragic, disastrous, unsatisfactory	satisfactory, positive, successful, favourable	unintended, unexpected		
AFTERMATH	immediate	sad, bloody, war				
SEQUEL	logical, immediate, inevitable, possible		long-waited, interesting			
BY-PRODUCT	unavoidable, inevitable	undesired, corrosion, harmful, unfortunate	intriguing, valuable, useful	unpredicted, unexpected, bizarre, surprising, incidental, accidental	important	
UPSHOT	certain	deleterious				
END-PRODUCT						

Table 5.16 Semantic sets of modifiers of the words in query

In Table 5.17, the symbol √ indicates that we could find collocates or co-occurrences in this semantic set while X means that no collocates or co-occurrences could be found in the semantic set. In addition percentages of the semantic sets out of all the instances of the query word are offered in the table, for example, for *RESULT*, about 10.7% out of 300 instances are categorized as having LOGIC semantic association.

Semantic Sets	Logic	Evaluation		Unexpectedness	Seriousness	Same /difference	Total
		Negative	Positive				
RESULT	√ 10.7%	√ 2.7%	√ 5.9%	X	X	√ 6.7%	26%
EFFECT	√ 6.4%	√ 8.4%	√	X	√ 3.5%	√ 0.7%	19%
CONSEQUENCE	√ 7.6%	√ 13.4%	X	√ 0.4%	√ 6.1%	√ 0.1%	27.5%
IMPACT	√ 6.8%	√ 2.8%	√	X	√ 14.1%	√ 5.4%	39.7%
FRUIT	X	X	X	X	X	X	0
OUTCOME	√ 13.6%	√ 1.3%	√ 8.7%	√ 0.8%	X	X	24.4%
AFTERMATH	√ 57.7%	√ 4.9%	X	X	X	X	63.6%
SEQUEL	√ 17.5%	√ 5%	X	X	X	X	22.5%
BY-PRODUCT	√ 14.8%	√ 10.2%	√ 0.7%	√	√ 10.2%	X	35.9%
UPSHOT	√ 11.1%	√ 11.1%	X	X	X	X	22.2%
END-PRODUCT	√ 15%	X	√ 35%	X	X	√ 10%	60%

Table 5.17 Distribution and percentage of semantic sets of modifiers of the words in query

The overlap in the semantic sets seems to support to some extent the view that these candidate words share similar meanings or senses. It can be seen from the table that all the words in query (except *FRUIT*) have the LOGIC association but with different proportions. To be specific, *AFTERMATH* has the top proportion (57.7%) and *IMPACT* has the lowest (only 6.8%) with the LOGIC association, with all the others in between, which suggests similarities and differences among the candidate words in terms of their associations with the semantic set LOGIC. In a similar way, *CONSEQUENCE*, *OUTCOME* and *BY-PRODUCT* share an association with the semantic set UNEXPECTEDNESS while others do not have this semantic association. Therefore, the overlapping of semantic associations of the words under investigation lead us to conclude that the candidate words have a degree of closeness or similarity and that the different strength of the associations indicate their distance or divergence. In other words we can demonstrate how some words are more synonymous than others but it is difficult to say whether two words have met the criteria of synonymy.

5.2.3.4. Colligation

5.2.3.4.1. grammatical distribution of the query words in the clause

Now that we have considered the collocation and semantic association of the words in query, we now turn to colligation. Hoey (2005, p. 44) points out 'a noun will always be part of some group or other word sequence and that group or word sequence will normally perform some function in a clause. One can therefore look at the distribution of any noun in terms of its occurrence within clause or group' (also see Sinclair, 1991). In this section I examine the distributions of the nouns under investigation in both clause and group with a view to comparing their distributions.

A random sample of 300 instances of each lemma was retrieved from the BNC (except for *SEQUEL, BY-PRODUCT, UPSHOT* and *END-PRODUCT* where the full sample was examined since the hits of these four words were 301, 254, 154 and 57 respectively), to see whether they occurred as part of the Subject, as part of the Object, as part of the Complement or as a part of a prepositional phrase functioning as Adjunct. Following Hoey (2005), I define Object 'as having a different referent from Subject unless it is filled by one of the self-reflexive pronouns such as *himself* and as characteristically following transitive verbs'. Hoey (2005) also points out:

> As anyone who attempts the grammatical analysis of authentic data knows, one encounters rather more cases where a correct analysis is problematic than one might anticipate on the basis of conveniently simple, made-up examples. It is not always possible to distinguish postmodification, particularly of an adjective, from a prepositional phrase functioning as Adjunct; Adjuncts and postmodifying prepositional phrases are not quite as neatly separable as one might imagine. Particles following a verb are another area where existing criteria do not always let one arrive at an intuitively satisfying analysis. (p. 46)

As BNC consists of 10% spoken data, which is often characterised by fragmental sentences or clauses, some instances had to be excluded from the data. Thus for *SEQUEL, BY-PRODUCT, UPSHOT* and *END-PRODUCT* I only have 292, 251, 154 and 57 instances for colligational analysis. Also following Hoey, I excluded those instances whose 'senses …were clearly separable and idiomatic uses that did not retain the word's priming function' (Hoey, 2005, p. 45).

The figures of instances and percentage of the grammatical position in the sample are given in Table 5.18. Taking *RESULT* as an example, 111 instances are found to appear as part of the Subject in the clause, which accounts for 27.7% of the sample of 300 instances. The highest and lowest proportions of each grammatical position are shown in bold and italics respectively. The category Others refers to part of headings or names of tables in texts. They usually appear as segments of a clause or sentence. They cannot really be described with a grammatical function, but to find out whether these query words have specific features, they are still included in the analysis.

The following findings deserve attention: Firstly, each word is primed with different proportional distributions in terms of grammatical position in the clause. For example, there is a clear positive colligation between *EFFECT* and the grammatical function of Object, as almost half of *EFFECT* occurs within Object. On the other hand, there is a negative colligation between *IMPACT* and the Complement function, with only 1.7% within Complement, and also *EFFECT* and the Complement function with a percentage of 2.7%.

	Part of subject	Part of object	Part of complement	Part of adjunct	Others	Total
RESULT	111 (37%)	76 (25.3%)	51 (17%)	42 (14%)	20 (6.7%)	300
EFFECT	83 (27.7%)	143 (47.7%)	8 (2.7%)	57 (19%)	9 (3%)	300
CONSEQUENCE	78 (26%)	104 (34.7%)	48 (16%)	68 (22.7%)	*2 (0.7%)*	300
IMPACT	60 (20%)	**156 (52%)**	*5 (1.7%)*	55 (18.3%)	**24 (8%)**	300
FRUIT	78 (26%)	128 (42.7%)	19 (6.3%)	65 (21.7%)	10 (3.3%)	300
OUTCOME	117 (39%)	79 (26.3%)	28 (9.3%)	67 (22.3%)	9 (3%)	300
AFTERMATH	*38 (12.7%)*	23 (7.7%)	10 (3.3%)	**207 (69%)**	22 (7.3%)	300
SEQUEL	106 (36.3%)	83 (28.4%)	35 (12%)	63 (21.6%)	5 (1.7%)	292
BY-PRODUCT	63 (25.1%)	33 (13.1%)	**83 (33.1%)**	72 (28.7%)	0	251
UPSHOT	**131 (85.2%)**	*6 (3.9%)*	4 (2.6%)	*13 (8.4%)*	0	154
END-PRODUCT	16 (28.1%)	14 (24.6%)	14 (24.6%)	13 (22.8%)	0	57

Table 5.18 A Comparison of the grammatical distributions of the candidate words in the clause

Secondly, among all the candidate words, *AFTERMATH* is the most negatively primed to occur with the function of Subject. To compensate, there is a positive colligation between *AFTERMATH* and the function of Adjunct. The data analysis shows that *AFTERMATH* occurs within Adjunct in almost 7 out of 10 cases. Examples are:

> That art was demonstrated with conspicuously different talents by British editors and journalists in the ***aftermath*** of the revelation that Jeffrey Archer, best-selling novelist and deputy chairman of the Conservative party, had paid a Shepherds Market street-walker £2,000 to leave the country.
>
> Fourteen officers, including Okar, and over 200 lower ranking soldiers were arrested in the immediate ***aftermath***.

Thirdly, among all the candidate words, *IMPACT* is the most negatively primed (1.7%) to occur with the function of Complement, but is the most positively primed (52%) with the function of Object. Examples are:

> Dicey's work has had a major and lasting ***impact***.
>
> … like other schools, trends in educational theory and practice have an immediate ***impact*** on them.

Next, there is a positive colligation between *UPSHOT* and the function of Subject. However, *UPSHOT* is negatively primed with other functions including Object, Complement and Adjunct:

> The ***upshot*** is that small areas of the boundary layer are turbulent.
>
> Anyway, the ***upshot*** was that he demanded there be a committee meeting this Thursday to work out a club strategy.

Finally, Table 5.18 shows that *RESULT* is ten times collocating as in Others than *CONSEQUENCE*, and that *IMPACT* is roughly four times in Others than *SEQUEL*. Compared with the other words in the set, *RESULT, IMPACT* and *AFTERMATH* occur in headings more often, with a percentage of 6.7%, 8% and 7.3% respectively. It seems that these three words tend to appear as part of headings or names of tables in texts. Here are some examples:

> THE ELECTION ***RESULTS***
>
> ***aftermath*** of Temple Mount killings
>
> An evaluation of the ***impacts*** and effectiveness of news about nature conservation

5.2.3.4.2. colligational priming when subject

Since most of the candidate synonyms occur as Subject at least a quarter of the time, it seems worth giving close attention to the details of the candidate words as part of Subject. Following Hoey (2005), I looked at the definiteness and indefiniteness of the candidate words. I used the same data as above and found that the numbers of each word functioning as part of Subject are as follows: *EFFECT* (84), *RESULT* (79), *IMPACT* (54), *CONSEQUENCE* (78), *FRUIT* (76), *OUTCOME* (117), *AFTERMATH*

(37), *SEQUEL* (106), *BY-PRODUCT* (63), *UPSHOT* (131) and *END-PRODUCT* (16). The full results are to be found in Tables 5.19, 5.20 and 5.21.

Except for *FRUIT* and *BY-PRODUCT*, all the words colligate with definiteness, of which *UPSHOT* is the most strongly primed with definiteness, accounting for 97.7% of the data, followed by *AFTERMATH* (91.9%), *END-PRODUCT* (87.5%), *IMPACT* (87%), *RESULT* (83.5%), *EFFECT* (81%), *CONSEQUENCE* (75.6%), *OUTCOME* (72.6%) and *SEQUEL* (58.5%).

Words in Query	Definite	Indefinite	Total
RESULT	66 (83.5%)	13 (16.5%)	79
EFFECT	68 (81%)	16 (19%)	84
CONSEQUENCE	59 (75.6%)	19 (24.4%)	78
IMPACT	47 (87%)	7 (13%)	54
FRUIT	29 (38.2%)	47 (61.8%)	76
OUTCOME	85 (72.6%)	32 (27.4%)	117
AFTERMATH	34 (91.9%)	3 (8.1%)	37
SEQUEL	62 (58.5%)	44 (41.5%)	106
BY-PRODUCT	21 (33.3%)	42 (66.7%)	63
UPSHOT	128 (97.7%)	3 (2.3%)	131
END-PRODUCT	14 (87.5%)	2 (12.5%)	16

Table 5.19 Definiteness and indefiniteness of the candidate words

There are three main ways in which a nominal group may be made definite: with the definite article, with a possessive expression and with a determiner (Hoey, 2005). A close examination of how definiteness is realised in the Subject nominal groups shows that *UPSHOT* is 100% primed to occur with the definite article '*the*', followed by *END-PRODUCT* (92.9%). Among all the words, *FRUIT* is least frequently primed with the definite article 'the', with a percentage of 72.4%. To compensate, 17.2% occur with possessive expressions and 10.3% with a determiner. In addition, *SEQUEL* (72.6%) and *FRUIT* (72.4%) are primed to occur with the definite article with almost the same percentage but *SEQUEL* colligates more often with possessive expressions (22.6%) than with determiners (4.8%). Furthermore, *IMPACT*, *CONSEQUENCE*, *AFTERMATH*, *BY-PRODUCT*, *UPSHOT* and *END-PRODUCT* do not colligate with the determiners at all, which could be categorised as negative colligations.

Definite	*the*	*possessive*	*this/this/these/those*	Total
RESULT	55 (83.3%)	4 (6%)	7 (10.6%)	66
EFFECT	56 (82.4%)	8 (11.8%)	4 (5.9%)	68
CONSEQUENCE	54 (91.5%)	5 (8.5%)	0	59
IMPACT	39 (83%)	8 (17%)	0	47
FRUIT	21 (72.4%)	5 (17.2%)	3 (10.3%)	29
OUTCOME	75 (88.2%)	2 (2.4%)	8 (9.4%)	85
AFTERMATH	29 (85.3%)	5 (14.7%)	0	34
SEQUEL	45 (72.6%)	14 (22.6%)	3 (4.8%)	62
BY-PRODUCT	19 (90.5%)	2 (9.5%)	0	21
UPSHOT	128 (100%)	0	0	128
END-PRODUCT	13 (92.9%)	1 (7.1%)	0	14

Table 5.20 The distribution of markers of definiteness across the candidate words

On the other hand, *BY-PRODUCT* is the most frequently primed with indefiniteness, accounting for almost 67% of the data; while *UPSHOT* is primed to avoid indefiniteness, with less than 3% of instances occurring in an indefinite expression. Table 5.21 shows the distributions of markers of indefiniteness across the candidate words.

Indefinite	*a/an*	*another*	*one*	*any*	(None)	Total
RESULT	1 (7.7%)	0	0	0	12 (92.3%)	13
EFFECT	3 (18.8%)	2 (12.5%)	1 (6.3%)	0	10 (62.5%)	16
CONSEQUENCE	4 (21.1%)	0	5 (26.3%)	1 (5.3%)	9 (47.4%)	19
IMPACT	3 (42.9%)	0	0	0	4 (57.1%)	7
FRUIT	2 (4.3%)	0	0	0	45 (95.7%)	47
OUTCOME	6 (18.8%)	0	4 (12.5%)	1 (3.1%)	21 (65.6%)	32
AFTERMATH	1 (33.3%)	0	0	0	2 (66.7%)	3
SEQUEL	25 (56.8%)	0	1 (2.3%)	1 (2.3%)	17 (38.6%)	44
BY-PRODUCT	25 (59.5%)	3 (7.1%)	4 (9.5%)	0	10 (23.8%)	42
UPSHOT	0	0	1 (33.3%)	0	2 (66.7%)	3
END-PRODUCT	1 (50%)	0	0	0	1 (50%)	2

Table 5.21 The distribution of markers of indefiniteness across the candidate words

5.2.3.4.3. grammatical distribution of the candidate words in the nominal group

We turn now to the grammatical preferences and aversions of the candidate words at the rank of the group or phrase, looking at whether they occur as the head of the nominal group in which it appears, as a premodifier or as part of the postmodification. Examples of the grammatical possibilities are:

> … then *the **effect** of this* on the computer system and its environment, in particular its effect on the working method of project team members, must be clearly identified.
>
> The inside of two large lorries had been converted into *a special **effects** roadshow* that allows visitors to get the sound, sight, taste, smell and feel of Guinness - and win prizes as they go.
>
> While there is *concern **over the long-term effects** of population losses from northern regions of Britain*, in some circles there appears to be even more anxiety about the failure of migration to produce a speedier matching of workers to jobs.

	Head of nominal group	Part of the postmodification of the nominal group	Premodifier of nominal group
RESULT	293 (97.7%)	5 (1.7%)	2 (0.7%)
EFFECT	280 (93.3%)	18 (6%)	2 (0.7%)
CONSEQUENCE	294 (98%)	6 (2%)	0
IMPACT	273 (91%)	15 (5%)	12 (4%)
FRUIT	232 (77.3%)	5 (1.7%)	63 (21%)
OUTCOME	286 (95.3%)	9 (3%)	5 (1.7%)
AFTERMATH	297 (99%)	3 (1%)	0
SEQUEL	278 (93%)	11(3.7%)	10 (3.3%)
BY-PRODUCT	237 (93.3%)	9 (3.5%)	8 (3.1%)
UPSHOT	153 (99.4%)	1 (0.6%)	0
END-PRODUCT	51 (89.5%)	6 (10.5%)	0

Table 5.22 Grammatical distributions of the candidate words in the nominal group

Again, I compare the frequencies of the candidate words in any of these three positions and Table 5.22 shows the result. First of all, all the nouns occur most frequently as heads of their own nominal groups. Secondly, *FRUIT* occurs least frequently as the head of a nominal group among all the words. To compensate, it is strongly primed to occur as premodifier of a nominal group. But this is a possible case of the effect of polysemy. As in *fruit salad, fruit cup* and *fruit ball, fruit* is used as premodifiers and also in literal use. This may suggest a possible method of identifying polysemous uses of *FRUIT*, and therefore the instances in which *FRUIT* functions as a premodifier should be eliminated and we

only focus on the instances of '*FRUIT* of…' structure. Thirdly, *EFFECT, RESULT, IMPACT, OUTCOME, SEQUEL* and *BY-PRODUCT* show a small tendency to occur as premodification. *CONSEQUENCE, AFTERMATH, UPSHOT* and *END-PRODUCT* show no tendency at all. Finally, among all the words, *END-PRODUCT* is most strongly primed to occur as part of the postmodification of the nominal group, with a percentage of 10.5%.

5.2.3.4.4. characteristic primings with respect to theme

As shown in the previous section, there are differences among the candidate words in terms of their positions in a clause, namely as part of the Subject, as part of the Object, as part of the Complement or as a part of a prepositional phrase functioning as Adjunct. Examination of the data likewise reveals their different status as Theme. Halliday (1994) defines Theme as 'the element which serves as the point of departure of the message; it is that with which the clause is concerned' (p. 37). Thus according to Halliday (1994), the Theme of a clause 'ends with the first constituent that is either participant, circumstance, or process' (p. 52) and Rheme is 'the remainder of the message' (p. 67), i.e. everything which is not Theme.

Building on Halliday's work on Theme, Berry (1995; 1996) argues that Theme does not necessarily refer to only the first ideational element in a clause and argued that Theme should be extended up to and including the Subject. The main difference between Halliday's model and Berry's model is shown in Table 5.23 with an example.

<table>
<tr><td></td><td>On a clear day,</td><td>you</td><td>can see forever</td></tr>
<tr><td>Halliday's model</td><td>Theme</td><td colspan="2">Rheme</td></tr>
<tr><td rowspan="2">Berry's model</td><td colspan="2">Theme</td><td>Rheme</td></tr>
<tr><td>Additional Theme</td><td>Basic Theme</td><td>Rheme</td></tr>
</table>

Table 5.23 Main difference between Halliday's and Berry's model of Theme and Rheme Analysis

In the current analysis, Theme and Subject may be the same in a clause, but they refer to different textual notions. For example:

> Already as a consequence of the war, half the children up to five years are short for their age due to malnutrition.

In this sentence, *half of the children up to five years* is the Subject. However, all of *Already as a consequence of the war, half the children up to five years* will be included in the Theme, according to the criteria proposed by Berry (1995; 1996), namely any textual material preceding the main verb.

	Theme	Rheme	Total
RESULT	105 (35%)	195 (65%)	300
EFFECT	89 (29.7%)	211 (70.3%)	300
CONSEQUENCE	100 (33.3%)	200 (66.7%)	300
IMPACT	61 (20.3%)	239 (79.7%)	300
FRUIT	69 (23%)	231 (77%)	300
OUTCOME	120 (40%)	180 (60%)	300
AFTERMATH	111 (37%)	189 (63%)	300
SEQUEL	91 (30.3%)	209 (69.7%)	300
BY-PRODUCT	71 (28%)	183 (72%)	254
UPSHOT	137 (89%)	17 (11%)	154
END-PRODUCT	17 (29.8%)	40 (70.2%)	57

Table 5.24 Distributions of the words in query as Theme and Rheme

The Subject Themes, also known as unmarked Themes, are not of special interest. Examination of the samples reveals that all the words in query (except *UPSHOT*) are part of Theme with a proportion of below 40%. However Almost 90% of instances of *UPSHOT* occur as Theme in the clauses in which it appears, among which about 70% of them are subjects (Table 5.24).

	Sentence-initial clauses	Non-sentence-initial clauses	All clauses
RESULT	73 (69.5%)	32 (30.5%)	105
EFFECT	56 (62.9%)	33 (37.1%)	89
CONSEQUENCE	61 (61%)	39 (39%)	100
IMPACT	42 (68.9%)	19 (31.1%)	61
FRUIT	49 (71%)	20 (29%)	69
OUTCOME	71 (59.2%)	49 (40.8%)	120
AFTERMATH	83 (74.8%)	28 (25.2%)	111
SEQUEL	67 (73.6%)	24 (26.4%)	91
BY-PRODUCT	52 (73.2%)	19 (26.8%)	71
UPSHOT	115 (83.9%)	22 (16.1%)	137
END-PRODUCT	13 (76.5%)	4 (23.5%)	17

Table 5.25 Distributions of the words in query in sentence-initial and non-sentence-initial clauses

When it comes to whether they are used in Sentence-initial or Non-sentence-initial clauses, the words in query have similar distributions (Table 5.25). However, almost 84% of instances of *UPSHOT* and only 59% of instances of *OUTCOME* appear in sentence-initial clauses, being the first and the last of all the words in query in this respect.

Table 5.26 and 5.27 demonstrate the proportions of initial Themes of the words in query when they function as Subjects, Adjuncts or Other clausal functions. In the sentence-initial clauses, *UPSHOT* is positively primed with Subjects being the highest proportion (93.9%) among all the words in query, but it is negatively primed with Adjuncts to appear in Thematised Adjuncts with the lowest proportion of 6.1% and it does not appear in other functions at all. *AFTERMATH* is negatively primed with Subject with a proportion of only 19.3%. To compensate, it is positively primed with function of Adjuncts with the highest proportion of 79.5% among all the query words.

	Subjects	Adjuncts	Other clausal functions	Total (sentence-initial clauses)
RESULT	40 (54.8%)	28 (38.4%)	5 (6.8%)	73
EFFECT	35 (62.5%)	15 (26.8%)	6 (10.7%)	56
CONSEQUENCE	36 (59%)	22 (36.1%)	3 (4.9%)	61
IMPACT	25 (59.5%)	9 (21.4%)	8 (19%)	42
FRUIT	39 (79.6%)	4 (8.2%)	6 (12.2%)	49
OUTCOME	57 (80.3%)	9 (12.7%)	5 (7%)	71
AFTERMATH	16 (19.3%)	66 (79.5%)	1 (1.2%)	83
SEQUEL	52 (77.6%)	12 (17.9%)	3 (4.5%)	67
BY-PRODUCT	33 (63.5%)	13 (25%)	6 (11.5%)	52
UPSHOT	108 (93.9%)	7 (6.1%)	0	115
END-PRODUCT	11 (84.6%)	2 (15.4%)	0	13

Table 5.26 Distribution of initial Themes in sentence-initial clauses

In non-sentence-initial clauses, *END-PRODUCT* only appears as the Subject (with a proportion of 100%). Next to it, *UPSHOT* is positively primed with Subject with the highest proportion of 95.5%. Again *AFTERMATH* is positively primed with function of Adjunct with the highest proportion of 67.9% among all the words in query.

	Subjects	Adjuncts	Other clausal functions	Total (non-sentence-initial clauses)
RESULT	23 (71.9%)	8 (25%)	1 (3.1%)	32
EFFECT	27 (81.8%)	3 (9.1%)	3 (9.1%)	33
CONSEQUENCE	27 (69.2%)	10 (25.6%)	2 (5.1%)	39
IMPACT	14 (73.7%)	3 (15.8%)	2 (10.5%)	19
FRUIT	18 (90%)	1 (5%)	1 (5%)	20
OUTCOME	44 (89.8%)	4 (8.2%)	1 (2%)	49
AFTERMATH	9 (32.1%)	**19 (67.9%)**	0	28
SEQUEL	22 (91.7%)	1 (4.2%)	1 (4.2%)	24
BY-PRODUCT	15 (78.9%)	1 (5.3%)	3 (15.8%)	19
UPSHOT	21 (95.5%)	1 (4.5%)	0	22
END-PRODUCT	4 (100%)	0	0	4

Table 5.27 Distribution of initial Themes in non-sentence-initial clauses

To conclude, the quantitative analysis has shown that these words in query share similarities and differences in terms of their collocational, semantic associational and colligational behaviour. Although the analyses offered cannot be regarded as precise enough to identify synonymy on their mown, they support the existence of a closeness or similarity of meanings or senses among the words in query.

5.3. Analysis of potentially synonymous items occurring in the same sentence

5.3.1. introduction and significance of the method

According to McEnery & Wilson (1996, p.76), ‘qualitative forms of analysis offer a rich and detailed perspective on the data’ and ‘qualitative studies enable one to discover which phenomena are likely to be genuine reflections of the behaviour of a language and which are merely chance occurrence’. A number of corpus linguists have emphasised the importance of combining both quantitative and qualitative analysis in corpus linguistics, as quantitative studies may give us information ‘in generalizations about language and language use’ (Kennedy, 1998) and qualitative analysis data are used for more than providing ‘real-life’ examples of particular phenomena. A couple of qualitative studies have contributed to the understanding of findings in quantitative analysis. For example, in the qualitative analysis of the corpus of Whitehouse briefings, Partington (2003) discusses the social situation, relationship between the press officers and press calls to understand the findings in quantitative analysis of the corpus, and he came up with more interesting findings.

In the previous section the findings of quantitative analysis have been offered. In this section, After looking at the general tendency of the divergences among these words in query, I will provide findings from the analysis of potentially synonymous items occurring in the same sentence. The analysis is more qualitative than quantitative in that the method could help us identify the analysis result that computers could not.

The analysis proposed here is very different from the commonly used corpus-linguistic method of looking at words in query in the concordance lines. This method does not concern itself with whether the two words appear within five-word span; it however is concerned with whether the two candidate synonyms appear in the same sentence.

5.3.2. procedure

The Sketch Engine was utilised to elicit those sentences in which two of the words in query appear at the same time. The purpose was to elicit a small set of corpus data with the candidate synonyms (for example *RESULT* and *CONSEQUENCE*) in the same context, to be specific, in the same sentence.

To avoid arbitrariness, two pairs of candidate synonyms were analysed. The pair *RESULT* and *CONSEQUENCE* was first looked at. To explore the relationship between these two candidate synonyms, a search of the lemma *RESULT* (as a noun) was made with any word form of *CONSEQUENCE* in the context of 15-word span on both sides (Figure 5.3). After removing those sentences where one of the two words was used as part of a verb and those where the co-occurrences crossed sentence boundaries, I looked at their semantic relationships in the sentences. Following the same method, an analysis with *RESULT* and *OUTCOME* was conducted to test whether the findings only exist for *RESULT* and *CONSEQUENCE*.

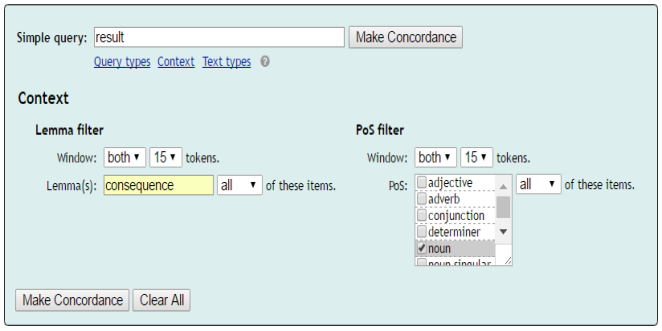


Figure 5.3 Snapshot of a search of the lemma *RESULT* (as a noun) with any word form of *CONSEQUENCE* in the context of 15-word span on both sides

5.3.3. findings

5.3.3.1. Co-hyponymy, synonymy and possible antonymy (oppositeness)

The analysis of *RESULT* and *CONSEQUENCE* seems to suggest three types of semantic relationships between the two words. First, in some cases the words are functioning as synonymous. As shown in examples 5.30 and 5.31, *RESULT* and *CONSEQUENCE* sometimes appear in parallel structure, meaning the same thing. The only differences among the following examples seems to be that both the words are used neutrally in 5.30; positively in 5.31 and negatively in 5.32.

Example 5.30:

Basins form as a *result* of downwarping of the crust, as a *consequence* of uplift of the surrounding region, or through a combination of both of these effects.

Example 5.31:

Partly as a *consequence* of incomes policy and more directly as a *result* of efforts at reform, the government became increasingly involved in trade unions.

Example 5.32:

As many as 50% of patients admitted to hospital following a successful resuscitation from out-of-hospital cardiac arrest will die before discharge, mainly as a *result* of cardiogenic shock or the *consequences* of lengthy anoxia.

Secondly, in some cases the words seem to set up a possible opposition. In example 5.33, the structure 'does not… but rather…' draws out a contrast between *as a result of healthy ageing* and *as a consequence of the development of atrophic changes of the gastric mucosa. The development of atrophic changes of the gastric mucosa* is definitely negative although *healthy ageing* is not absolutely positive. In this case, it seems that *RESULT* is used positively in contrast to *CONSEQUENCE* used negatively. Again in example 5.34, *result of deliberate government intention* and *the unintended (and embarrassing) consequence of minor regulatory change* again seem to pose a contrast. In 5.35, *primarily a result of deliberate and purposive employer labour strategy* and *more an unintended consequence of technological advance* are also in a possible opposition, linked by *rather than*.

Example 5.33:

This work shows that gastric acid secretion does not decline as a *result* of healthy ageing, but rather as a *consequence* of the development of atrophic changes of the gastric mucosa.

Example 5.34:

Rather than being the *result* of deliberate government intention, this reflected the unintended (and embarrassing) *conseqeunce* of minor regulatory change.

Example 5.35:

With this line of argument, however, it is possible to exaggerate the extent to which the hierarchical division of labour -- fostering sectionalism among workers -- was primarily a *result* of deliberate and purposive employer labour strategy, rather than being more an unintended *consequence* of technological advance.

Finally in some cases the two words are superordinate and subordinate, as in the following example: *consequence* is an *unintended result*.

Example 5.36:

This (almost certainly unintended) *result* is a *consequence* of the writer relying on other people to state ideas rather than trying to understand and restate them in her or his own voice.

To test whether this situation only exists with *RESULT* and *CONSEQUENCE*, an analysis with *RESULT* and *OUTCOME* was conducted following the same method. Again three types of semantic relationships between the two words seem to emerge.

First, the two words are again functioning as synonyms. In example 5.37, *RESULT* and *OUTCOME* appear in parallel structure, meaning the same thing. Also in 5.38, the two words are synonymous, both being positive.

Example 5.37:

To encourage developers to write more modular software that can take advantage of systems that support Threads - a Thread can be any part of an application or programme that is not dependent on the *result* or *outcome* of another (that can be treated as a task in its own right) - Posix has a committee working on a Threads application programming interface standard.

Example 5.38:

Issues of economics and safety had been exhaustively debated. Whatever the anti-nuclear movement might have said about the way the *result* was achieved, the CEGB had got the *outcome* it wanted.

Secondly, the two words seem to be in a possible opposition, as in examples 5.39 and 5.40 where *RESULT* and *OUTCOME* are linked by *not…but…*, which poses a contrast.

> Example 5.39:
>
> It is thought that the Criminal Law Revision Committee, the Report of which formed the basis of the Theft Act, would not have wanted such a *result* but would have preferred the *outcome* to be theft.
>
> Example 5.40:
>
> Achieving this situation is not a random *OUTCOME* but is the *RESULT* of adopting a proactive process-based approach to all aspects of team working and the provision of appropriate training.

Finally the two words appear to be in a relationship of superordinate and subordinate, as in 5.41: *outcomes* may be *the unintended results*.

> Example 5.41:
>
> Contemporary studies of both policy making and policy implementation suggest that we need to give attention to some very complex relationships between the mixed goals of those able to influence policies and the varied consequences of their interventions. *Outcomes* may be the *unintended results* of policy inputs.

Note that in example 5.34, the two words in query are ten words apart; therefore they can never be picked up as collocation if we only look at them within the usual five-word-span of concordance lines. Mostly the corpus approach pays more attention to L1, L2 or R1, R2 collocates; this method however takes a new approach, being interested in the sentences in which the two words appear rather than in the concordance lines.

5.3.3.2. metaphor and synonymy

As shown before, the word *FRUIT* does not share primings with other query words in terms of their collocations, semantic associations and colligations. This may be due to the small number of instances of the word being used in a metaphorical sense in the corpus, so an analysis of *FRUIT* in its

metaphorical sense seems appropriate. The literal sense of *FRUIT* is incapable of being synonymous to the other words in query; therefore only one polysemous sense of *FRUIT* is discussed here.

It however is not easy to elicit the instances of *FRUIT* used in a metaphorical sense from the corpus. Macmillan English Dictionary provides two phrases which may be related to metaphorical use of the word *FRUIT*:

> **bear fruit** 1. to have a successful result: *Our policies must be given time to bear fruit*. 2. If a tree or plant bears fruit, it produces fruit
>
> **the fruit/fruits of sth** the good results that you get from something such as hard work: *The book is the fruit of a collaboration between several groups*. ♦ **the fruits of your labour** *Retirement is a time to relax and enjoy the fruits of your labour*.

As a way of looking at the collocational and colligational behaviours of *FRUIT* when used in a metaphorical sense, 139 instances of '*bear fruit(s)*' and 485 instances of '*fruit(s) of …*' were elicited from the BNC corpus.

It was my prediction that the metaphorical sense of *FRUIT* is mostly used in plural form and this prediction was checked against my data. The analysis of instances of '*bear fruit(s)*' shows that altogether 9.4% are used in a literal sense and 90.6% in a metaphorical sense, and that 125 out of 126 instances (almost 99%) of the metaphorical sense are in singular form.

Literal (13; 9.4%)		Metaphorical (126; 90.6%)		Total
Simple form	Plural form	Simple form	Plural form	
10 (7.2%)	3 (2.2%)	125 (89.9%)	1 (0.7%)	139

Table 5.28 Proportions of word forms of *FRUIT* in '*bear fruit(s)*' in the literal and metaphorical senses

Some examples follow:

> With more efficient use of the existing electronic information resources in our Town Halls this is an approach that could bear much *fruit*.
>
> A phone call to Maria bore no *fruit*.

Then 485 instances of '*fruit(s) of …*' were analysed, among which it was found that the singular form appears in 172 instances and the plural form in 313 instances. Of all the instances, 83.5% (405) are used in metaphorical sense, of which 139 (28.7%) are in singular form and 266 (54.8%) in plural form. In addition, nine instances are used in both literal and metaphorical senses.

Literal (71; 14.6%)		Metaphorical (405; 83.5%)		In-between (9; 2%)		Total
Singular	Plural	Singular	Plural	Singular	Plural	
33 (7%)	38 (8%)	139 (28.7%)	266 (54.8%)	/	9 (2%)	485

Table 5.29 Proportions of *fruit(s)* in '*fruit(s) of …*' in their literal and metaphorical senses

Some examples of metaphorical sense both in singular and plural forms are the following:

> Success was the *fruit* of some three years' strenuous work.
>
> In recent years much has been drawn from other denominations and a much wider choice of hymns and music is one of the *fruits* of ecumenism.

The following example is an example of an instance that is categorised as in between literal and metaphorical senses.

> The Halloween game of bobbing apples - catching apples in a tub of water with your mouth - probably originated as an ancient harvest rite, possibly again in honour of the Roman goddess Pomona. In Teutonic mythology, heaven was likened to a vale of apple trees tended by the goddess Idun. The apples were the *fruits* of perpetual youth and gave the gods immortality.

To sum up, the results to some extent contradict my prediction that *FRUIT* in metaphorical sense is usually in the plural form. As shown above, almost 90% of instances of *FRUIT* in the phrase '*bear fruit(s)*' are used in a metaphorical sense and also in singular form. Over 83.5% (405 instances) of the structure '*fruit(s) of…*' have a metaphorical sense, of which 139 (34.3%) instances are in singular form and 266 (65.7%) in plural form.

5.4. Corpus evidence to explain findings in the psycholinguistic experiment

Both the corpus analysis and contextualised analysis show that the candidate words share similarities in terms of their collocation, semantic association and colligation, which seems to suggest we could justify their closeness in meaning. However it is difficult to decide whether two words are synonyms or not, and sometimes in some contexts the candidate words may be considered as forming other types of semantic relations (for example, co-hyponymy or possible antonymy). This result seems to support

the findings in the psycholinguistic experiment reported in the previous chapter. This section will draw upon more corpus analysis to provide possible explanations of the findings in the experiment.

5.4.1. directionality of synonymy

In the experiment reported in the previous chapter, an interesting observation seems to relate to the directionality of the offered synonyms. For example, 17 participants wrote *agree* as synonymous with *accept*, but only 6 considered *accept* to be a synonym of *agree*. In other words, *agree* is provided as a synonym of *accept* more often than vice versa. To explore the causes, I turn to corpus data for possible explanations.

It turns out that *AGREE* is more frequent than *ACCEPT*, with 24,061 (214.28 per million) instances and 20,320 (181.00 per million) instances respectively in BNC, which might suggest that words with lower frequency may more readily elicit words with high frequency as synonyms. Take another pair as an example: seven participants considered *consequence* to be synonymous with *by-product* while only two provided *by-product* as a synonym of *consequence*. A corpus search in the BNC shows 7,763 (69.20 per million) occurrences of *CONSEQUENCE* and 254 (2.26 per million) instances of *BY-PRODUCT*. A simple explanation would be that people encounter the high frequency words more often and thus find them easier to recall.

Words in query	Frequency in BNC
AGREE	24,061 (214.28 per million)
ACCEPT	20,320 (181.00 per million)
CONSEQUENCE	7,763 (69.20 per million)
BY-PRODUCT	254 (2.26 per million)

Table 5.30 Frequency of *AGREE, ACEEPT, CONSEQUENCE* and *BY-PRODUCT* in BNC

The directionality of synonymy may also be related to other matters such as the numbers of senses of the words in question and the range of genre/text types in which the words are used. However, the experiment reported did not permit the gathering of evidence about the possible effects of these factors and further exploration of this topic is recommended.

4.5.3. possible scale of strength of synonymy

In addition, some pairs of synonyms are found to be 'more synonymous' than other pairs in the experiment. For instance, for *agree*, 14 people offered *concur* as synonym, 6 provided *accept* and only 2 wrote down *approve*. This seems to suggest that *agree* and *concur* are more synonymous than *agree* and *accept*, which in turn are more synonymous than *agree* and *approve*, which may point to something like a scale of synonymy, shown as follows:

Absolute Synonymy → Near-Synonymy → Non-Synonymy

Along this scale, we seem to have the following situation:

agree & concur > agree & accept > agree & approve

The experiment result seems to show there is a scale of synonymy among the candidate synonymous words provided by the participants. However, how do we measure this synonymy? As previously mentioned, when participants are asked to provide synonyms for a word, they tend to offer the most frequent candidate synonym first. But if we had looked only at the frequencies of the words *AGREE, ACCEPT, APPROVE* and *COCNUR* in the BNC (See Table 5.31), we would have come up with the scale *AGREE* & *ACCEPT* > *AGREE* & *APPROVE* > *AGREE* & *CONCUR*. However, the experiment has shown some divergence, in other words, frequency of words is not the only factor which may influence our decision on synonyms.

	Frequency in BNC
AGREE	24,061 (214.28 per million)
ACCEPT	20,320 (181.00 per million)
APPROVE	5,241 (46.72 per million)
CONCUR	247 (2.2 per million)

Table 5.31 Frequency of *AGREE, ACCEPT, APPROVE* and *CONCUR* in BNC

According to Hoey's (2005) theory of Lexical Priming, 'lexis is completely and systematically structured'. The likely priming effects, or the priming potential of repeated encounters with a word in its context, shared by synonymous items**,** reflect the close similarity of sense, but 'synonyms differ in respect of the way they are primed for collocation, semantic associations, colligations and pragmatic associations' (Hoey, 2005). By analyzing the synonymous pair *RESULT* and *CONSEQUENCE,* Hoey (2005) demonstrated that the two words share similar collocates and semantic associations but differ in the strength of distributions. Thus I looked at the collocational and colligational behaviours of these

four words (*AGREE, ACCEPT, APPROVE* and *CONCUR*). I sampled 300 instances of each word and looked at the collocations, semantic associations and colligations for each of these words.

The following are the adverb co-occurrences which modify the words in query:

AGREE	unhappily, readily, absently, voluntarily, previously, immediately, momentarily, eventually (4), initially, generally, completely, wholly, mutually, nationally, broadly, abjectly, apparently, personally (2), verbally.
ACCEPT	reluctantly (3), obviously, generally (4), subsequently, passively, finally (2), widely (6), willingly (2), successfully, tacitly, promptly, recently, uncritically, readily, commonly, sportingly
APPROVE	basically, exactly, wholly (2), partially, apparently, thoroughly, overwhelmingly (2), formally (7), finally (3), unanimously (2), previously, subsequently, presumably, wholeheartedly, implicitly
CONCUR	reluctantly, cheerfully, readily (3), wholeheartedly (2), overwhelmingly, provisionally, roughly, largely, completely, wholly, entirely (3), fully (2), thoroughly, unanimously, broadly, emphatically, feelingly, lifelessly, undoubtedly, overtly, apparently, obviously, certainly, duly, strongly

It can be seen that there are some overlaps of the adverb collocates among the four words, *e.g.* *AGREE* and *ACCEPT* share *readily* and *generally.* This seems to support the view that they share similar meanings or senses, in other words, that they are near-synonymous but differ in the number of shared adverbs and in the frequency of the adverbs shared (e.g. *AGREE* and *COCNUR* share 4 adverbs but none occurs more than once). I have briefly identified six semantic sets for *AGREE* and *COCNUR* (more details of which will be given in the next chapter). They are:

1. co-occurring adverbs such as *reluctantly*, *unhappily*, *cheerfully*, *voluntarily*, *readily*, *wholeheartedly*, *passively* and *willing* express (UN)WILLINGNESS:

2. co-occurring adverbs such as *finally*, *previously*, *eventually*, *provisionally*, *immediately*, *promptly*, *momentarily*, *recently*, *initially* and *subsequently* denote a semantic set of STAGE/TIME.

3. co-occurring adverbs including *generally*, *entirely*, *completely*, *wholly*, *roughly*, *largely*, *fully*, *thoroughly*, *widely* and *partially* are classified as belonging to a semantic set of EXTENT.

4. collocates such as *unanimously*, *mutually*, *nationally*, *broadly* and *overwhelmingly* belong to the semantic set RANGE.

5. collocates such as *respectfully*, *emphatically*, *apparently*, *obviously*, *personally*, *certainly*, *undoubtedly*, *overtly* and *presumably* belong to the semantic set of ATTITUDE or STANCE) (of the speakers or writers).

6. collocates *feelingly* and *lifelessly* and co-occurring adverbs *unhappily*, *abjectly*, *cheerfully* and *reluctantly* form a semantic set of EMOTIONS (of the subjects of the sentences).

Note all the four words under consideration share the first four semantic sets but with different collocates in the sets (Table 5.32).

		AGREE frequency	*CONCUR* frequency	*ACCEPT* frequency	*APPROVE* frequency
(UN) WILLINGNESS	*reluctantly*	0	1	1	0
	unhappily	1	0	0	0
	cheerfully	0	1	0	0
	voluntarily	1	0	0	0
	readily	1	3	1	0
	wholeheartedly	0	2	0	1
	passively	0	0	1	0
	willingly	0	0	1	0
STAGE/TIME	*finally*	0	0	2	3
	previously	1	0	0	1
	eventually	4	0	0	0
	provisionally	0	1	0	0
	immediately	1	0	0	0
	promptly	0	0	1	0
	momentarily	1	0	0	0
	recently	0	0	1	0
	initially	1	0	0	0
	subsequently	0	0	1	1
EXTENT	*generally*	1	0	4	0
	entirely	0	3	0	0
	completely	1	1	0	0
	wholly	1	1	0	2
	roughly	0	1	0	0
	largely	0	1	0	0
	fully	0	2	0	0
	thoroughly	0	1	0	1
	widely	0	0	6	0
	partially	0	0	0	1
RANGE /SCOPE	*unanimously*	0	1	0	2
	mutually	1	0	0	0
	nationally	1	0	0	0
	broadly	1	1	0	0
	overwhelmingly	0	1	0	2

ATTITUDE or STANCE (of the speakers or writers)	*respectfully*	0	0	0	0
	emphatically	0	1	0	0
	apparently	1	1	0	1
	obviously,	0	1	1	0
	personally	2	0	0	0
	certainly	0	1	0	0
	undoubtedly	0	1	0	0
	overtly	0	1	0	0
	presumably	0	0	0	1
EMOTIONS (of the subjects of the sentences)	*feelingly*	0	1	0	0
	lifelessly	0	1	0	0

Table 5.32 Differences in collocates and semantic associations of *AGREE, CONCUR, ACCEPT* and *APPROVE*

In addition to considering the collocations and semantic associations of the four words, I looked at their colligational behaviours. The following Table 5.33 shows the different proportional distribution of their word forms.

AGREE	agree	agrees	agreeing	agreed
300	113 (37.7%)	12 (4%)	10 (3.3%)	165 (55%)
ACCEPT	accept	accepts	accepting	accepted
300	151 (50.3%)	16 (5.3%)	26 (8.7%)	107 (35.7%)
APPROVE	approve	approves	approving	approved
300	54 (18%)	8 (2.7%)	8 (2.7%)	230 (76.7%)
CONCUR	concur	concurs	concurring	concurred
246	116 (47.2%)	24 (8.1%)	8 (3.3%)	98 (39.8%)

Table 5.33 Proportional distribution of word forms of the lemmas

The analysis seems to show that *AGREE & CONNCUR* are more closely synonymous than *agree & accept* and *agree & approve.* The result shows that 37.7% of *AGREE* and 47.2% of *CONCUR* are used in the infinitive from while the percentages for *ACCEPT* and *APPROVE* are 50.3% and 18% respectively. In other words, *AGREE* and *COCNUR* are more similar to each other in terms of their uses in the infinitive form. As for the use of past participle, the percentages for *accepted*, *concurred*, *agreed* and *approved* are 35.7%, 39.8%, 55% and 76.7%. which again shows that *AGREE* and *CONCNUR* are closer than the pairs *AGREE* & *ACCEPT*, and *AGREE* & *APPROVE*.

As shown above, the candidate synonyms provided by the participants in the experiment seem to suggest that there is a scale of synonymy, in other words, *AGREE* & *CONCUR* is more synonymous than *AGREE* & *ACCEPT*, and in turn both pairs are more synonymous than *AGREE* & *APPROVE*. Given Bawcom's point (2010) that frequency is one of the most important factors in determining synonyms, it might have been thought that the frequency list would suggest how synonyms are stored in people's minds (Chapter 4). The experiment result however seemed to contradict the frequency list of the four in the BNC. I therefore conducted a corpus analysis of the four words and found that the similarities and differences between the four words in terms of collocations, semantic associations and colligations provided more reliable information about how similar and different these words are. The corpus analysis seemed to support the idea of a scale of synonymy for the four words, that is, *AGREE* & *CONCUR* > *AGREE* & *ACCEPT* > *AGREE* & *APPROVE*, which was consistent with what was found in the experiment. To sum up, along with frequency, the collocational and colligational behaviours of candidate synonyms has been found to play a vital role in determining which words are deemed the most closely synonymous.

5.5. Conclusions of the chapter

This chapter has achieved the following goals. Firstly, a corpus-driven analysis of eleven potentially synonymous words in English has shown that a corpus approach is capable of demonstrating similarities and differences among these putative synonyms. By using the categories in lexical priming such as collocation, semantic association and colligation, this chapter has measured the strength of synonymy among these words. In other words, the strength of synonymy among the eleven candidate synonyms has been shown in their primings with respect to their different proportions in collocations, semantic associations and colligations. Secondly the corpus approach has proved to be effective in identifying potential synonyms, but cannot determine whether these words are indeed synonyms. Because there is a scale of similarity, we could only (can only claim that) some words are very synonymous while others are slightly synonymous. The analysis however seems to show that there is no effective corpus-based method to distinguish semantic relations such as synonymy and co-hyponymy, which might suggest the distinction between these semantic relations may be blurred.

Despite the challenges, this study has explored the possibility of using a corpus-driven approach to identifying similarities and differences among potential synonyms. The degree of equivalence or similarity in meanings of candidate words can be measured and computed, even allowing for the fact that statistical measurements could be improved, and these measurements might be used to quantify the semantic distance between apparently synonymous forms.

In addition this chapter has also helped answer the third research question stated in section 5.2.1, that is, can the results of corpus analysis help explain the findings in the psycholinguistic experiment? If we can make an analogy between mental concordance and corpus concordance, we may offer the explanation that via encounters with different language data in various contexts people's minds may be primed to group words in certain ways, for example, words frequently appearing in similar contexts and co-texts may share closeness in meaning, which therefore might be considered as synonyms.

To conclude, this chapter has found that the distinction between synonymy and other semantic relations can be blurred and there is no neat way to distinguish synonymy from co-hyponymy or metonymy as it concerns some issues including distinction between synonymy of words and synonymy of senses, and also statistical distortion of the polysemy senses. The notion of synonymy in English is valid but synonymy is a very complicated language phenomenon and the concept needs to be modified referring to the categories utilised in lexical priming. In the next two chapters Chinese corpus data will be investigated to see whether the findings concerning synonyms derived from the analysis of Chinese data are consistent with the findings concerning English synonyms derived from the same kind of analysis of English data. In addition, if we find synonymy can be described in the same way in languages which have no family relationship, it will also be investigated whether the corpus-linguistic categories used by Lexical Priming enable us to identify similarities and differences between candidate synonyms in both English and Chinese.

Chapter 6 The applicability of Lexical Priming to Chinese Synonyms: a case study comparing a pair of potential English synonyms with a pair of potential Chinese synonyms of equivalent meaning

6.1. Introduction to the chapter

The previous chapters have partly answered the first two research questions of the thesis by focusing on synonymy in English. In Chapter 4, a psycholinguistic experiment was carried out to explore the psychological reality of synonymy. The result showed that people consider words as synonymous in a way that is not as we had expected. For the same prompt words, people provided various candidate words as synonyms, which indicated that it is difficult to pin down words as exact synonyms or not, and which led us to wonder whether a corpus analysis of natural language data might come up with similar findings. Therefore in Chapter 5 a group of assumed English synonyms was analysed and it was found that a corpus approach could help us identify the similarities and differences among these words. We showed the strength of similarities among the candidate synonyms by using the categories employed by lexical priming. However we could only say that some words are very synonymous and others are synonymous to a certain degree, but we could not decide whether two words are synonyms or non-synonyms as sometimes the boundary between synonymy and other word relations could be fuzzy. Due to their both being on the scale of similarity there is no easy way to distinguish between synonymy and co-hyponymy; therefore we concluded that the concept of synonymy in English is valid but needs modifications.

After looking at synonymy in English, this chapter and the next chapter will focus on Chinese synonymy. Chinese and English are typologically different languages; to compare them, we need a framework that permits their comparability. Hoey's Lexical Priming seems to provide a useful framework. Hoey and Shao (2015) have demonstrated that the psychological and linguistic claims of Lexical Priming theory are not culture or language-specific. As preliminary observations on the applicability of Lexical Priming theory to Chinese have been presented in that paper, the aim of this chapter is test the claims of lexical priming concerning synonyms on Chinese data. Hoey (2005) claims 'synonyms differ in respect of the way they are primed for collocation, semantic associations, colligations and pragmatic associations' and supported the claim with an analysis of the English synonymous pair *result* and *consequence*. Therefore this chapter explores whether Chinese near-synonyms are primed differently in terms of their collocations, colligations, semantic associations and pragmatic associations. A pair of Chinese near synonyms 同意 (*tóng yì*) and 赞同 (*zàn tóng*) has been chosen for the analysis. However, in order to make the comparison more explicit between Chinese and English, an analysis of the English equivalents *AGREE* and *CONCUR* is also presented.

All the chosen words are epistemic verbs, which often include both descriptive and performative meanings (cf. Traugott, 1989; Nuyts, 2001). Take *AGREE* for example. One of the senses describes a

shared page that people are having the same opinion while another sense is related to the performative sense of the word, as when someone agrees to do something, he/she is making a promise. The focus of the study however is not on the distinction of polysemous senses, but on the measurement of the shared sense of the words in query; therefore to avoid distortion of attempts to measure the strength of similarities among my chosen synonymous verbs, only shared sense (i.e. *to have the same opinion*) is included in the study and the data relating to the other senses (for example, *to promise to do something*) have to be excluded from the current analysis.

6.2. Purpose and Research Questions

This chapter aims to investigate how far Hoey's hypothesis regarding synonymy is supported by Chinese data, to be specific, to explore the primings of Chinese near-synonyms in terms of their collocations, colligations, semantic associations and pragmatic associations. In other words, this chapter is to test whether the strength of similarities between words can be measured by the categories applied in Lexical Priming and used in the previous chapter. Since the hypothesis has been provisionally supported on the basis of examination of English synonymous nouns (*consequence* and *result*, the pair Hoey used to illustrate the theory, 2005) and in the previous chapter, the focus in this chapter is on verbs. Therefore my research questions are:

(1) Are members of English near-synonymous verb pairs or sets also primed differently for collocation, semantic association, colligation and pragmatic association?

(2) If we find that pairs of English near-synonymous verbs are primed differently for collocation, semantic association, colligation and pragmatic association, is the same true of Chinese near-synonymous verbs?

(3) If the claim is supported for Chinese, are there any similarities and differences between the pairs or sets of Chinese and English near-synonyms in terms of collocation, semantic association, colligation and pragmatic association?

6.3. Methodology: data and analysis tool

To tackle these questions, the British National Corpus (BNC) was analyzed for English near-synonyms and zhTenTen11 for Chinese near-synonyms using the Sketch Engine (Kilgariff, 2003). The Sketch Engine is one of the few language analysis tools which can analyse both English and Chinese data. It offers various applications such as word sketch and sketch difference in addition to the expected concordance and word list functions. As mentioned before, a word sketch is a one-page summary of the word's grammatical and collocational behaviour. It shows the word's collocates categorised by grammatical relations such as words that serve as an object of the verb, words that serve as a subject of the verb, words that modify the word etc. Word sketch difference is used to compare and contrast two words by analysing their collocations and by displaying the collocates divided into categories based on

grammatical relations (Kilgarriff, 2003). In the previous chapter I used word sketch to look at a group of English candidate synonyms and in this chapter it seems more appropriate to utilize sketch difference to analyse the candidate synonymous pairs.

ZhTenTen11 is a web crawling Chinese corpus in simplified characters (versus traditional characters, currently only used in Taiwan and partly in Hong Kong for historical and political reasons) collected in 2011. It contains 2,106,661,021 characters. The web corpus can be invaluable when a large quantity of data is needed for the study of language; it however has some specific problems. One drawback of web crawling data is that it is difficult to trace the source of data and therefore almost impossible to know the genre or text type of each instance. In spite of this, the findings of the current study are unlikely to be seriously affected since we are more interested in the general characteristics of synonymy, although we do have to recognise that we will not be able to pick up on whether genre and context affect synonyms in Chinese.

The English near equivalents *AGREE* and *CONCUR* (in capitalisation to refer to the lemma of the word) and their Chinese equivalents 同意 *(tóng yì)* and 赞同 *(zàn tóng)* were chosen for the following reasons. Firstly, as noted earlier, previous studies of synonymous items within the framework of Lexical Priming have concentrated on nouns; therefore this chapter will focus on verbs to explore whether the lexical priming claims are also applicable to other grammatical categories. Secondly, the move from near-synonymous nouns to near-synonymous verbs is not just to look at synonymy across the different lexical categories. The synonymous nouns *result* and *consequence* are also lexical signals for discourse markers. The analysis of synonymous verbs *AGREE* and *CONCUR* will therefore also provide observations on reporting words, which are another type of lexis with discourse functions.

There are 23,123 instances (206.1 per million) of *AGREE* and only 247 instances (2.2 per million) of *CONCUR* in the BNC corpus, which strongly suggests that *AGREE* is much more frequent than *CONCUR* in English. No bottom-up analysis of the senses of the English words was conducted, but rather the corpus-based dictionaries were consulted, as a source of informed but independent judgements.

All the dictionaries show that our chosen verbs are polysemous although the number and nature of senses the dictionaries provide differ from one another. According to Hoey (2005), 'the collocations, semantic associations and colligations a word is primed for will systematically differentiate its polysemous senses' (p. 81). As his 'drinking problem' hypotheses have been tested by a couple of studies (Hoey, 2005; Pace-Sigge, 2015), the focus here in not on providing more evidence. What seems to be related to the current study of synonymy is that if one word has several senses, their collocations, semantic associations and colligations with different senses may influence the statistical significance of attempts to distinguish synonyms. To avoid distortion of attempts to measure the strength of similarities among my chosen synonymous verbs, this paper only focuses on a sense that the two words share and

which is present in all three dictionaries, viz. 'to have the same opinion or to reach an agreement', and the data relating to the other senses (for example, *to be consistent with*) have been excluded from the current analysis. This left me 205 instances of *CONCUR*, and therefore a sample of 205 instances of *AGREE* with the same sense was retrieved from the corpus for the convenience of statistical comparison.

In my Chinese corpus, there are 152,083 instances (72.2 per million) of 同意 *(tóng yì)* and 13,706 instances (6.5 per million) of 赞同 *(zàn tóng)*, which shows that, as with the English pair, 同意 *(tóng yì)* is much more frequent than 赞同 *(zàn tóng)* in the corpus. Since there was no corpus-based Chinese dictionary available, I conducted bottom-up analysis to decide the senses of 同意 *(tóng yì)* and 赞同 *(zàn tóng)*. I identified three senses of 同意 *(tóng yì),* which can be glossed in English as follows:

(1) to have a similar opinion;

(2) to say one will do something as suggested;

(3) to grant official permission.

My analysis of the data for 赞同 *(zàn tóng)* suggests that the word has two main senses:

(1) to have a similar opinion; and

(2) to approve of, accept.

Since the current study is only concerned with apparently synonymous items, senses 2 and 3 for 同意 *(tóng yì)* and sense 2 for 赞同 *(zàn tóng)* were eliminated from the analysis in case the results were distorted by those senses of the words. A sample of 250 instances of 同意 *(tóng yì)* was compared with a sample of 250 instances of 赞同 *(zàn tóng)* for analysis of Chinese near-synonyms.

Note that deciding whether the instances should be included in the analysis is not as straightforward as one might think. There are two reasons for this. Firstly the traditional description of grammatical categories may not always meet our needs. Sinclair (1991) examined the functions of the word '*of*' and pointed out that *of* – considered to be a preposition in traditional grammar – in most cases does not function the way that prepositions are supposed to. He argues that 'prepositions are principally involved in combining with following nouns to produce prepositional phrases which function as adjuncts in clauses' and 'this is not anything like the main role of *of*, which combines with preceding nouns to produce elaboration of the nominal group', for instance, '*the back of the van, a small bottle of brandy'* (Sinclair, 1991). He consequently excludes *of* from the class of prepositions.

This phenomenon is not rare in English and very common in Chinese. Since there is no inflectional variation in Chinese, it is even more difficult than in English to decide the part of speech of words. For example:

Example 6.1

居然　　没有　　找到　　一 个　赞同　　的　　声音 …. . .

Jū rán méi yǒu zhǎo dào yí gè zàn tóng de shēng yīn.
Surprisingly have not find one concurring PAR voice…
Surprisingly (we) have not found one concurring voice/ one voice which concurs.

Whether 赞同 *(zàn tóng)* here should be considered as *concurring* (adj.) or *concur* (verb) is a tough decision. In fact there is no reason why 赞同 *(zàn tóng)* cannot be considered as a noun since nouns have the feature of functioning as modifiers. This phenomenon supports Hoey's (2005) assumption that 'lexis is chosen first or at least earlier' and that choice of words is not the result of slotting into given grammatical categories. Wherever difficulties in assigning a grammatical category occurred, the instances in question were removed from the data since they are not what the current study is concerned with.

Secondly, it has to be noted that both in English and Chinese the senses of a word overlap to some extent and in some cases it is hard to say that one sense absolutely dominates in the sentence. For example, the senses of *AGREE* in the following two sentences from the Macmillan English Dictionary overlap to some extent.

Example 6.2:

We all *agree* that we should celebrate this event.
('have the same opinion as someone else')

Example 6.3:

We need to *agree* on a date for our next meeting.
('say that you will do something that someone else wants or suggests')

In example 6.2, it can be argued that we are also articulating that we will celebrate the event (i.e. will do something) since this agreement has to be expressed verbally. And presumably, this is a suggestion from somebody (the one who wants or proposes celebrating the event), since there is a very low probability that we would reach an agreement if the topic has never been suggested or mentioned by someone. Example 6.3 can also be explained in terms of the need to reach an agreement that on a certain date we'll have the next meeting. In this sense, *AGREE* can also mean 'to have the same opinion as others'.

With the Chinese data, a similar situation occurs:

Example 6.4:

全体	合伙人	同意	转让……
Quán tǐ	hé huǒ rén	tóng yì	zhuǎn ràng……
All	co-partner	agree	transfer……

All the co-partners have agreed to transfer…

We could interpret the meaning of example 6.4 as 'all the co-partners have agreed to transfer…' or 'all the co-partners have reached an agreement that they would transfer…', in which case, we would say senses of the word 同意 *(tóng yì)* again overlap to some extent.

Each of the samples of candidate synonymous usages in both English and Chinese were analysed. To test the statistical significance of the differences found between the synonyms, Rayson's log-likelihood was used then based on the proportional distribution of senses in the ZhTenTen11 corpus. Note that in the following part of the paper, all the statistics in log-likelihood are based on the figures in the original data rather than on the sample.

6.4. Results and Analysis

I present the analysis of the English data first.

6.4.1. analysis of the English data

6.4.1.1. collocation and semantic association

The first part of my analysis concerns collocation. The analysis shows that *AGREE* and *CONCUR* share similar collocates but that the proportional distribution of the collocates varies between the two words. The difference in the distribution of the two words with respect to prepositional collocates *with, in* and *on* can be seen in Table 6.1. The two words show similarities in distribution with respect to prepositions *with* and *on*, but have marked differences in distribution with the proposition *in*. In Table 6.1 three sets of statistics are presented: first, the number of the concordance lines in which the lemmas co-occur with the prepositions; second, the percentage of co-occurrences of the lemma and the three prepositions in the 205 instances of my sample. For example, the information in the first cell of the matrix (63; 30.7%) shows that 63 instances of *AGREE* co-occur with the preposition *with*, and that accounts for 30.7% of the 205 instances of the lemma *AGREE* in the sample. Lastly, log-likelihood scores are presented in the table, which indicates the collocation strength of the collocates and the words in query. Note that the higher the score, the more evidence we have that the differences between the words are greater (Hardie, 2012).

	with	*in*	*on*
AGREE (205)	63; 30.7%	0	2; 1.0%
CONCUR (205)	85; 41.5%	29; 14.1%	2; 1.0%
Log-likelihood	**6.83**	**257.03**	0

Table 6.1 Instances and proportions of collocates (prepositions) with *AGREE* and *CONCUR*

There are also differences between the two words in respect of their co-occurrence with pronouns (Table

6.2). The word *AGREE* seems to prefer to collocate with *I* (47 instances, 22.9%) and *you* (12 instances, 5.9%) while *CONCUR* has only 22 co-occurrences with *I* (10.7%) and 6 with *you* (2.9%). On the other hand, *CONCUR* favours the pronoun *we* more than *AGREE* does (8.8% and 3.9% respectively). This divergence between the two synonymous verbs in terms of strength of collocation will be talked about again with regard to their colligations.

	I	*We*	*you*
AGREE (205)	47 (22.9%)	8 (3.9%)	12 (5.9%)
CONCUR (205)	22 (10.7%)	18 (8.8%)	6 (2.9%)
LL	**16.44**	**9.03**	3.65

Table 6.2 Instances and proportions of collocates (pronouns) with *AGREE* and *CONCUR*

There are some co-occurring words, e.g. adverbs such as *wholeheartedly,* e*ntirely* and *unanimously,* that intuitively seem to be collocates. The numbers of the instances, however, are so few that it is impossible to do more than identify them as worthy of further exploration with a larger database. And larger database does provide us with evidence for semantic associations.

Firstly, modifiers are used to indicate the degree of (UN) WILLINGNESS of the person/people who agrees/agree. The modifiers in my data include adverbials such as *reluctantly, readily* and *wholeheartedly.* Examples are:

> chiefs said: 'The association **wholeheartedly agrees** with the views expressed by Mr Adair. It is…'
> … would be secure, and only after a long argument **agreed** very **reluctantly** that he could come back on duty

Secondly, words such as *finally, previously* and *eventually* refer to the STAGE/TIME of reaching an agreement. For example:

> … to meet the higher self-financing targets they had **previously agreed** with the Ministry.
> 'Well, thank heaven you **finally agree** with me!' was all his comment.

The third semantic set, which in my data includes *generally* and *entirely*, comprises words used to indicate the EXTENT of agreement.

> Historians **generally agree** that the outcome was favourable to the outcome was favourable to the government.
> Not all analyses **agree entirely** with this conclusion.

There also appears to be a fourth set, exemplified in my data by *unanimously, mutually* and *nationally,* which are used to show the RANGE/SCOPE of people arriving at an agreement.

The summit **unanimously agreed** that a market of 320 million people would improve growth prospects and trade.
NeXT and van Cuylenburg **mutually agreed** that the restructured 200-person company no longer requires both a chief executive and a president and chief operating officer.

Lastly, there is evidence of a semantic set where there is an expression of EMOTION, but this seems only to occur with *CONCUR*.

'No, indeed!' Theda **concurred feelingly**.
'It happened, and now it's over,' she **concurred lifelessly**.

It has been found that there are shared semantic associations between the two words with only one exception but that the selections made of words from these semantic sets may nevertheless differ (Table 6.3). Note that the figures in Table 6.3 are raw frequencies along with the significance of the collocates co-occurring with the lemmas. The statistical association measure used in Sketch Engine (http://www.sketchengine.co.uk) is logDice, a score which 'has a reasonable interpretation, scales well on a different corpus size, is stable on subcorpora, and the values are in reasonable range' (Rychlý, 2008).

Semantic sets	collocates	*AGREE* (frequency; significance)	*CONCUR* (frequency; significance)
(UN) WILLINGNESS	*reluctantly*	20; 4.8	1; 5.9
	readily	35; 5.5	3; 5.9
	wholeheartedly.	25; 7.9	1; 7.7
STAGE/TIME	*finally*	58; 5.7	0
	previously	3; 1.7	0
	eventually	20; 4.4	0
	provisionally	5; 2.8	1; 7.5
EXTENT	*generally*	124; 6.9	0
	entirely	112; 8.8	3; 4.3
RANGE/SCOPE	*unanimously*	40; 8.5	30; 8.0
	mutually	1; 6.0	0
	nationally	28; 8.0	0
EMOTIONS (of the subjects of the sentences)	*feelingly*	0	1; 9.3
	lifelessly	0	1; 9.3

Table 6.3 Differences in collocates and semantic associations of *AGREE* and *CONCUR*

To sum up, the English synonymous verbs *AGREE* and *CONCUR* are alike in sharing similar collocates and semantic sets but differ in the distributions and proportion of collocates from the same semantic set.

6.4.1.2. colligation

6.4.1.2.1. word forms

As noted before, colligation refers to "the grammatical position and function a word tends to prefer in or avoid" (Hoey, 2005). In order to test whether near synonymous verbs differ in the way they are primed for different grammatical patterns and functions, first I looked at the distribution of the word forms of the two English synonyms. The preliminary analysis shows that out of the 205 instances of *AGREE*, there are 109 instances of *agree*, 9 instances of *agrees*, 4 instances of *agreeing* and 87 instances of *agreed* and the data of 205 instances of *CONCUR* yield 98 instances of *concur*, 17 of *concurs*, 4 of *concurring* and 86 of *concurred* (Table 6.4).

AGREE	agree	agrees	agreeing	agreed
205	105 (51.2%)	9 (4.4%)	4 (2%)	87 (42.4%)
CONCUR	concur	concurs	concurring	concurred
205	98 (47.3%)	17 (7.8%)	4 (2%)	86 (42%)
Log-likelihood	0.47	**5.53**	0.00	0.01

Table 6.4 Proportional distribution of word forms of the lemmas

Even though the two words do not show any difference in proportion of occurrence for *–ing* and *–ed* forms, the most noticeable point, however, is that *CONCUR* occurs in the third singular form with a proportion of 7.8%, almost twice that of *agrees* (4.4%).

Based on this, I also looked at the distributions of *agreed* and *concurred* across simple past, perfect and passive (Table 6.5). Analysis shows that out of 87 instances of *agreed*, 58 (66.7%) are used for simple past tense, 6 (6.9%) for perfect aspect and 23 (26.4%) for passive voice. In 86 hits of *concurred*, 74 (86%) are in simple past tense and 8 (9.3%) for perfect aspect, but only 4 (4.7%) are used in passive voice.

	simple past	perfect tense	passive voice
agreed (87 instances)	58 (66.7%)	6 (6.9%)	23 (26.4%)
concurred (86 instances)	74 (86%)	8 (9.3%)	4 (4.7%)
Log-likelihood	2.13	0.31	**14.56**

Table 6.5 Distribution of *agreed* and *concurred* between simple past, perfect and passive

It has been widely acknowledged that different forms of a lemma behave differently and have different meanings (Renouf, 1986; Sinclair, 1991; Stubbs, 1996; Tognini-Bonelli, 2001). The current analysis seems to strongly support that synonyms share similarities as regards colligation with particular verb forms and functions but also differ in some ways in terms of the grammatical categories they prefer.

6.4.1.2.2. subjects

We now turn to the Subjects of the two verbs. It can be seen from Table 6.6 that *AGREE* and *CONCUR* differ in the degree to which they associate with different grammatical categories. Whereas the difference in occurrence with singular nouns is not large (19.5% vs. 27.3%), we can see a clear difference as far as the plural nouns are concerned. The proportion of *CONCUR* co-occurring with plural nouns is almost twice as high as that of *AGREE*. Furthermore, *AGREE* co-occurs with first person pronoun singular form (22.9%) more than twice as often as does *CONCUR* (10.7%). As far as the second person pronouns and other pronouns such as *few*, *many* and etc. are concerned, the numbers of instances and proportions are higher (more than twice) with *AGREE* than with *CONCUR*. On the other hand, as regard to the first person pronoun plural form, *CONCUR* occurs with more instances and in higher proportions than *AGREE*.

	Animate Subjects								Inanimate Subjects
	Nouns		Personal Pronouns					Other Pronouns (few, many, etc.)	
	Singular	Plural	First Personal		Second (you)	Third Personal			
			Singular (I)	Plural (we)		Singular (he/she)	Plural (they)		
AGREE (205 instances)	40 (19.5%)	26 (12.7%)	47 (22.9%)	8 (3.9%)	12 (5.9%)	17 (8.3%)	9 (4.4%)	14 (6.8%)	32 (15.6%)
CONCUR (205 instances)	56 (27.3%)	45 (22%)	22 (10.7%)	18 (8.8%)	6 (2.9%)	16 (7.8%)	7 (3.4%)	6 (2.9%)	29 (14.1%)
LL	5.60	11.20	16.44	9.03	3.65	0.06	0.48	5.78	0.29

Table 6.6 Proportions of different types of Subjects occurring with *AGREE* and *CONCUR*

There seems to be no significant difference concerning the frequency of inanimate subjects with the two verbs. The sub-categorization of these inanimate subjects, however, appears to indicate much divergence. Table 6.7 shows how *AGREE* and *CONCUR* differ in the detail of their co-occurrence with inanimate subjects. Over 85% of the instances of *CONCUR* appear with a committee or a particular organization as the Subject in the active voice whereas the figure for *AGREE* is below 30%. When the sentences are in the passive voice and an inanimate Subject is drawn from IDEAS & OPINIONS semantic association, the proportion for *AGREE* is almost five times as much as that for *CONCUR*. It is also worth noting that 12 instances (37.5%) of AGREE occur in the structure *it is agreed that…,* but no instance of *CONCUR*, however, appears in the parallel structure, which suggests a significant difference between the two verbs in terms of this grammatical pattern.

	Inanimate Subjects			
	Committee, organization, etc. (in active voice)	Ideas, opinions (in passive voice)	It is + adj. + to + agree/concur…	It is + agreed /concurred + that …
AGREE (32 instances)	9 (28.1%)	10 (31.3%)	1 (3.1%)	12 (37.5%)
CONCUR (29 instances)	25 (86.2%)	2 (6.9%)	2 (6.9%)	0
LL	**18.73**	**9.49**	0.76	**23.86**

Table 6.7 Proportions of the inanimate subjects of *AGREE* and *CONCUR*

What seems interesting is the differences between *AGREE* and *CONCUR* in terms of their subjectivity/subjectification. Look at the following example:

> HMIs had already expressed a view that there was a prima facie case for the proposal, and ***the Council agreed*** that it would mean setting up a music board or at least an exploratory committee - and the DES music panel had offered suggestions of possible members.
> ***I agree*** wholeheartedly with his complaints and have been equally infuriated when missing wickets falling `live'.

There is a difference between '*the Council agreed*' and '*I agree*'. In the former case, it points out a matter of fact in a piece of information given by the speaker or writer. However, when the first personal pronoun is used, there is an overwhelming purpose of maintaining the interaction between the speaker and listener. Even if in writing, the purpose is not to point out the fact, but actually report some internal or subjective fact.

The analysis of *AGREE* and *CONCUR* has shown the difference in terms of their subjectivity/subjectification. Out of 205 instances, 22.9% of *AGREE* co-collocate with first personal pronoun singular form ('*I*'), which is more than twice as many as *COCNUR* (10.7%). While collocating with inanimate subjects such as *committee* or *organization*, *AGREE* has a percentage of 28.1% while *CONCUR* 86.2%. This suggests that *AGREE* is used in relation to subjectivity/subjectification more frequently than *COCNUR*.

6.4.1.2.3. objects

As a next stage the Objects of the two verbs were looked at. Following Quirk et al. (1985), I am choosing to treat that-clauses as objects rather than as subordinate/dependent clauses. Instances in passive voice

and also those in the patterns *it is agreed that…* and *it is* + adjective + *to* + *agree/ concur (that)…* were of course eliminated from the analysis and this therefore left me with 182 instances of *AGREE* and 201 of *CONCUR* for further analysis.

The biggest difference between the two verbs in terms of their Objects lies in those *that*-clauses used after the verbs. In total 31.8% of instances of *AGREE* are followed by a *that*-clause (with or without the use of *that*) whereas only 8.5% of *CONCUR* are used in this structure (Table 6.8).

	Proposition	That clause		Direct Speech	No Objects
		(with *that*)	(without *that*)		
AGREE (182 instances)	66 (36.3%)	49 (26.9%)	9 (4.9%)	15 (8.2%)	43 (23.6%)
CONCUR (201 instances)	118 (58.7%)	16 (8.0%)	1 (0.5%)	14 (7.0%)	52 (25.9%)
LL	**32.62**	**29.93**	**11.53**	0.07	1.74

Table 6.8 Instances and proportions of objects of *AGREE* and *CONCUR*

As shown in Table 6.9, even with the same preposition *with*, what follows after the preposition shows divergence; for instance, out of 66 instances in which *AGREE* is followed by the preposition *with*, 6 (9.1%) are in the pattern '*AGREE with* sb. + *that* clause' whereas only 4 instances (3.4%) for *CONCUR* in the similar (the equivalent) structure. On the other hand, only 3 instances (4.5%) are used in the pattern '*AGREE with* + noun (e.g. fact, view) + *that* clause' while 11 instances (9.3%) for *CONCUR* are used in this structure. No instances were found of the preposition *in* after *AGREE* but 25 (21.2%) instances of *CONCUR* were identified used with *in.*

			AGREE (66 instances)	*CONCUR* (118 instances)	*LL*
Prepositions	*with*	sb./sth.	52 (78.8%)	66 (55.9%)	3.43
		+ *what* clause	2 (3.0%)	4 (3.4%)	1.52
		sb. + *that* clause	6 (9.1%)	4 (3.4%)	0.75
		noun (*fact, view*, etc.) + that clause	3 (4.5%)	11 (9.3%)	**12.34**
	in	sth.	/	25 (21.2%)	**221.58**
		+ *what* clause	/	1 (0.8%)	**8.86**
		noun (fact,..) + *that* clause	/	3 (2.5%)	**26.59**
	on	sth.	2 (3.0%)	1 (0.8%)	0.61
		+ *what* clause	/	1 (0.8%)	**8.86**
	over	+ *whether* clause	1 (1.5%)	/	1.99

Table 6.9 instances and proportions of objects with different prepositions of *AGREE* and *CONCUR*

6.4.1.2.4. adjuncts

As been mentioned before, there are some co-occurring adverbs, such as *wholeheartedly,* e*ntirely* and *unanimously,* that intuitively seem to be collocates. Further exploration with a larger database led us to find more collocates to form associations of semantic sets. So to test my hypothesis that these collocates may also form colligation, I utilized the Word Sketch to look at all the collocates as modifiers and additional adverb co-occurrences were found. In my data, 19 out of 205 instances of *AGREE* and 31 out of 205 instances of *CONCUR* co-occur with this type of adverb, which Greenbaum (1969, 1996) classifies as 'adjuncts', adverbs used to show manner, place, frequency, degree of intensity, etc. In this function they typically modify a constituent of a clause such as the verb or the predicative adjective. Table 6.10 shows the adjunct adverbs which co-occur with the synonymous pair and the number of occurrences in my sample.

AGREE	*normally(1), entirely(3), finally(1), tartly(1), perfectly(1), cautiously(1), eventually(2), generally(1), completely(1), absently(1), merely(1), personally(1), wholeheartedly(1), broadly(1), unconditionally(1)* and *further(1)*
CONCUR	*entirely(3), aggressively(2), further(1), duly(1), broadly(1), reluctantly(1), provisionally(1), emphatically(1), lifelessly(1), completely(1), apparently(1), fully(1), readily(3), wholly(1), strongly(2), obviously(1), largely(1), overwhelmingly(1), thoroughly(1), wholeheartedly(1), unanimously(1), duly(1), fully(1), undoubtedly(1)* and *certainly(1)*

Table 6.10 Adverb co-occurrences of *AGREE* and *CONCUR*

6.4.1.3. pragmatic associations

In lexical priming, Hoey (2005) points out:

> Just as a word or word sequence may be primed for semantic association, so it may be primed pragmatically as well. Pragmatic association occurs when a word or word sequence is associated with a set of features that all serve the same or similar pragmatic function (e.g. indicating vagueness, uncertainty). (p.26)

This section presents the analysis result of pragmatic associations.

6.4.1.3.1. expressing speaker/writer's attitude

It needs to be pointed out that the boundaries between semantic association and pragmatic association sometimes are blurred. For example, the words which comprises adverbials to express the speaker/writer's own ATTITUDE can form a semantic set and can be also used to serve a pragmatic function, namely, to express speaker/writer's attitude. These words include *kindly* and *respectfully*. Look the following example:

> I refer to our conversation last week at which you **kindly agreed**…
> Here again I **respectfully agree** with the observations made by Lord Donaldson M.R.

6.4.1.3.2. negation

Negation was the type of pragmatic association then concentrated on when the pragmatic associations

of the near-synonymous verbs *AGREE* and *CONCUR* were explored. It was found that the total instances and proportions of *AGREE* and *CONCUR* in conjunction with negation show almost no difference (7.3% vs. 7.8%) (see Table 6.11). On the other hand, of the 15 lines of negation with *AGREE*, all but one (93.3%) took the form of an expression defined as a broad negative in the Collins COBUILD English Grammar (Sinclair et al) and only one line (6.7%) used the word *fail*, which is an instance of an 'implied negative' (Quirk et al, 1985). On the other hand, out of 16 instances of *CONCUR*, 11 (68.8%) lines were identified as broad negatives and 5 (31.2%) as 'implied negatives' where items such as *difficult, refuse* and *refusal* were found in the concordances. Obviously the data are too few to draw confident conclusions, but they suggest a possible difference to be investigated with a larger set of data.

	Negation (total)	Negated modal verbs	Implied negatives
AGREE (205 instances)	15 (7.3%)	14 (93.3%)	1 (6.7%)
CONCUR (205 instances)	16 (7.8%)	11 (68.8%)	5 (31.2%)
LL	0.06	0.69	**7.91**

Table 6.11 instances and proportions of negation with *AGREE* and *CONCUR*

6.4.1.3.3. elicitation or confirmation of opinions

As has been mentioned before, *AGREE* and *CONCUR* are divergent in respect of their association with the second personal pronouns, with proportions of 5.9% and 2.9% respectively, which may be discussed again here in connection with the function of eliciting or confirming opinions. Among the 12 instances of *you* + *AGREE*, only one was used for eliciting opinions whereas 4 out of 6 lines of *you* + *CONCUR* were functioning as eliciting (2 lines) or confirming (2 lines) opinions. Again the data are sparse, but 8.3% versus 66.6% suggests a difference worthy of fuller investigation.

6.4.2. analysis of the Chinese data

6.4.2.1. collocation and semantic association

After presenting the analysis of English candidate synonymous pair *AGREE* and *CONCUR*, we now move to their Chinese equivalents 同意 *(tóng yì)* and 赞同 *(zàn tóng)*. As with the English data, the first part of the analysis concerns collocation and semantic association. Table 6.12 shows the instances and proportions of personal pronoun collocates with 同意 *(tóng yì)* and 赞同 *(zàn tóng)*. It can be seen from the table that the two words share collocates with regard to personal pronouns, but they are divergent in the distributions depending of the collocate. There are no big differences between the two words when the first personal pronouns are considered, and both the words favour singular (22% vs. 19.6%) rather than plural form (2.4% vs. 2.0%). This seem to suggest that 同意 (*tóng yì*) and 赞同 (*zàn tóng*) are not very different in their subjectivity/subjectification. Nonetheless, whereas 同意 *(tóng*

yì) tends to co-occur with the second personal pronoun singular form 你 *(nǐ)* and 您 *(nín)* (a 'respect' form, similar to that associated with *tu* and *vous* in French), the word 赞同 *(zàn tóng)* seems to favor third personal pronouns, both in singular and plural forms.

	Personal Pronouns					
	First Personal		Second Personal		Third Personal	
	Singular 我 *(wǒ)*	Plural 我们 *(wǒ men)*	Singular 你 *(nǐ)*	Singular (respect form) 您*(nín)*	Singular 他/她 *(tā/ tā)*	Plural 他们 *(tā men)*
同意*(tóng yì)*	55 (22%)	6 (2.4%)	8 (3.2%)	5 (2.0%)	7 (2.8%)	3 (1.2%)
赞同*(zàn tóng)*	49 (19.6%)	5 (2.0%)	2 (0.8%)	2 (0.8%)	19 (7.6%)	7 (2.8%)
LL	**22.78**	5.92	**230.62**	**81.84**	**420.71**	**118.61**

Table 6. 12 Instances and proportions of collocates with 同意 *(tóng yì) vs.*赞同 *(zàn tóng)*

Apart from the pronouns, analysis of the concordance lines also reveals collocates such as 双方 *(shuāng fāng)*(both sides), 一致 *(yí zhì)*(unanimously), 完全 *(wán quán)*(entirely/completely).

As was shown in the previous chapter, if we look at the word in a sentence rather than a concordance line, we pick up some findings which cannot otherwise be found. When a search for collocations for 赞同 *(zàn tóng)* was made using a five word span, no evidence was found of co-occurrence with 双方 *(shuāng fāng).* However, when the word span was extended to ten, four instances were retrieved in the corpus. In most studies of collocation in English, the word span is limited within L3 and R3 or L5 and R5; the need for a wider span in Chinese however is not occasional because of its particular features of syntactic structure; therefore when dealing with the Chinese data, a new methodology is proposed here. The collocation of the Chinese words in query is searched for with different word spans of 5, 10 and 15 tokens on both sides respectively to see whether we could obtain different findings from those with only L5 and R5 word span. With the search for collocations with 同意 *(tóng yì)* and 赞同 *(zàn tóng)*, different word spans yielded different results. Again differences in the number of both instances and strength of co-occurrence can be seen in Table 6.13.

		同意 *(tóng yì)*	赞同 *(zàn tóng)*
双方 *(shuāng fāng)* (both sides)	within 5 words on both sides	32	0
	within 10 words on both sides	39	4
	within 15 words on both sides	46	5
一致 *(yí zhì)* (unanimously)	within 5 words on both sides	31	3
	within 10 words on both sides	38	3

	within 15 words on both sides	49	3
完全 *(wán quán)* (completely or entirely)	within 5 words on both sides	11	17
	within 10 words on both sides	11	19
	within 15 words on both sides	14	24

Table 6. 13 Collocates of 同意 *(tóng yì) and* 赞同 *(zàn tóng)* within n-words on both sides

As have been mentioned, the need for a wider span is not occasional because of particular features of syntactic structure in Chinese. For example, in the following sentence:

Example 6.5:

双方 经过 长时间 讨论 ， 最后 同意 …
Shuāng fāng jīng guò cháng shí jiā tǎo lùn, zuìhoù tóng yì …
Both paties (TYPO) through long time discuss, finally agree…
After a long time discussion, both parties agreed…

Due to loose sentence structure in Chinese, the adjuncts 经过长时间讨论 (*jīng guò cháng shí jiā tǎo lùn*)(after a long time discussion) and 最后 (*zuì hoù*)(finally), can perfectly appear between the subject 双方 (*shuāng fāng*)(both parties) and the predicate 同意 (*tóng yì*) (agree), thus causing the phenomenon that the words in collocation may appear at a distance from each other. My previous analysis of Chinese data has led me to propose the notion of 'remote collocation' in comparison with the commonly used term collocation in corpus linguistics, which refers to the concurrence of word within usually 5 word spans. In fact, Hoey (2014) also argues for 'cohesive collocation' which he explains as 'words that occur in the local textual environment of the word under investigation but beyond the five word span' in his analysis of English data. Our studies were independent of each other and on typologically different languages, so it is of some interest that we have arrived at very similar conclusions.

After expanding the word span from five to fifteen tokens, I found more collocates and based on these, it was possible to identify three semantic sets that can be categorized as associating with the two words. Firstly, there is an association with the AGENTS involved in reaching an agreement, such as: 双方 *(shuāng fāng)(both sides),* 全体 *(quán tǐ)(all the members)* and 大家 *(dà jiā) (all the members)*. Secondly, there is a semantic association with IDEAS and OPINIONS, including, e.g. 意见 *(yì jiàn) (idea),* 观点 *(guān diǎn)(opinion)* and 论点 *(lùn diǎn) (argument).* Finally, there is a semantic association with the DEGREE/SCALE of agreement, including 完全 *(wán quán) (completely),* 部分 *(bù fēn)(partially)* and 基本 *(jī běn) (fundamentally).* Table 6.14 demonstrates the differences in semantic associations between the two verbs with respect to these associations.

		同意*(tóng yì)*	赞同*(zàn tóng)*
AGENTS	双方*(shuāng fāng)* (both sides)	46 (18.4%)	5 (2%)
	全体*(quán tǐ)* (all the members)	2 (0.8%)	0
	大家*(dà jiā)* (all the members)	15 (6%)	6 (2.4%)
IDEAS & OPINIONS	意见*(yì jiàn)* (idea)	31 (12.4%)	21 (8.4%)
	观点*(guān diǎn)* (opinion)	32 (12.8%)	46 (18.4%)
	论点*(lùn diǎn)* (argument)	2 (0.8)	1 (0.4%)
DEGREE/SCALE	完全*(wán quán)* (completely)	14 (5.6%)	24 (9.6%)
	部分*(bù fēn)* (partially)	3 (1.2%)	7 (2.8%)
	基本*(jīběn)* (fundamentally).	8 (3.2%)	7 (2.8%)

Table 6.14 Semantic associations of 同意 *(tóng yì)* and 赞同 *(zàn tóng)*

All in all, analysis of both the candidate English synonyms *AGREE* and *CONCUR* and their Chinese equivalents 同意 *(tóng yì)* and 赞同 *(zàn tóng)* has indicated how we may be primed with regard to collocations and semantic associations. As noted before, the current emphasis on the word span within 5 tokens on both sides in the definition of collocation needs more consideration.

6.4.2.2. Colligation

Since Chinese verbs do not change forms according to subject or tense, there is no need to consider the word forms with the Chinese data, so we look only at the co-occurrence of Subject and Object with the synonymous pair 同意 *(tóng yì)* and 赞同 *(zàn tóng)* for their colligational behaviours.

6.4.2.2.1 Subjects

As was the case with *AGREE* and *CONCUR*, the analysis of 同意 *(tóng yì)* and 赞同 *(zàn tóng)* shows different preferences with respect to different Subjects (see Table 6.15). The association with different personal pronouns was discussed when we looked at the collocation and semantic associations, which will not be repeated here. The two words 同意 *(tóng yì)* and 赞同 *(zàn tóng)* however also differ greatly with regard to the strength of their association with inanimate Subjects, namely 22.8% and 6% respectively.

	Animate Subjects								Inanimate Subjects	No Subjects
	Nouns	Personal Pronouns						Other Pronoun 大家 *(dà jiā)*		
		First Personal		Second Personal		Third Personal				
		Singular 我 *(wǒ)*	Plural 我们 *(wǒ men)*	Singular 你*(nǐ)*	Singular (respect form) 您*(nín)*	Singular 他/她 *(tā/ tā)*	Plural 他们 *(tā men)*			
同意*(tóng yì)*	90 (36%)	55 (22%)	6 (2.4%)	8 (3.2%)	5 (2.0%)	7 (2.8%)	3 (1.2%)	5 (2.0%)	57 (22.8%)	14 (5.6%)
赞同*(zàn tóng)*	141 (56.4%)	49 (19.6%)	5 (2.0%)	2 (0.8%)	2 (0.8%)	19 (7.6%)	7 (2.8%)	3 (1.2%)	15 (6%)	7 (2.8%)
LL	**786.18**	**22.78**	5.92	**230.62**	**81.84**	**420.71**	**118.61**	**32.07**	**1567.37**	**148.90**

Table 6. 15 Proportions of the subjects of 同意 *(tóng yì)* and 赞同 *(zàn tóng)*

A close examination of the sub-categorization of these inanimate Subjects also shows a clear divergence between the two words. Out of 57 instances in which inanimate nouns function as the Subjects, 27 (47.4%) of the concordance lines of 同意 *(tóng yì)* co-occur with 双方 *(shuāng fāng)* (both sides) whereas there is no instance of this expression with 赞同 *(zàn tóng)*. Moreover, 同意 *(tóng yì)* favours inanimate Subjects such as 专家组 *(zhuān jiā zǔ)*(expert team), 委员会 *(wěi 161uan huì)*(committee) and 董事会 *(dǒng shì huì)*(board of directors) with 8, 5 and 3 instances respectively, accounting for 28.1% of the data. On the other hand, different inanimate Subjects were used with 赞同 *(zàn tóng)* and except for three instances of 本网 *(běn wǎng)*(this website), only one instance was found of each inanimate subject co-occurring with 赞同 *(zàn tóng)*.

6.4.2.2.2. objects

The analysis of co-occurring Objects in the Chinese data is more challenging due to the fact that grammatical categorizations in China do not closely match those of English categories. The Objects in my corpus were roughly classified into three types: noun phrase, verb phrase, and clause, with a fourth group being 'no Object'. The distinction between verb phrase and clause in Chinese is much more difficult to make than it is in English because Chinese is a non-inflectional language and no forms of verbs can be found to indicate different grammatical structures. I decided to separate clauses from verb phrases whenever the clause could function as a separate sentence.

In Table 15, we can see significant differences between the two words in their association with noun phrase, verb phrase and clause as Objects. Apparently 同意 *(tóng yì)* favours verb phrases much more

than 赞同 *(zàn tóng)*, with proportions of 22.4% and 2.4% respectively. As far as noun phrases are concerned, more instances were found co-occurring with 赞同 *(zàn tóng)* than 同意 *(tóng yì)*. In addition, 赞同 *(zàn tóng)* tends to be used more frequently in the inverted structure, namely with the Object before the verb predicate, for instance:

Example 6.6:

他 的 观 点， 大家 都 赞同。
Tā de guān diǎn, dà jiā dōu zàn tong.
His opinion, everybody all concur.
Everybody concurs with his opinion.

In total 26.4% of 赞同 (zàn tóng) were used in this inverted structure while only 5.6% of 同意 (tóng yì) were found in the data with the same pattern. This finding can also be interpreted in terms of THEME/RHEME as a textual colligation.

Another divergence found between this candidate pair of synonyms lies in the fact that 同意 *(tóng yì)* appears to be used more in complicated sentence structureS, by which I mean the objects of the word 同意 (tóng yì) are complex sentences with multiple subjects and predicates. Altogether 27 instances of 同意 (tóng yì) and only 8 of 赞同 *(zàn tóng)* were identified in this structure, out of which 同意 *(tóng yì)* prefers non-direct speech while 赞同 *(zàn tóng)* favours direct speech with quotation marks.

	Object Noun Phrase		Object Verb Phrase	Object Clause		No Object
	(before verb)	(after verb)		Direct speech	Non-direct speech	
同意*(tóng yì)*	14 (5.6%)	104 (41.6%)	56 (22.4%)	1 (0.4%)	26 (10.4%)	49 (19.6%)
赞同*(zàn tóng)*	66 (26.4%)	116 (46.4%)	16 (6.4%)	6 (2.4%)	2 (0.8%)	44 (17.6%)
LL	**2820.32**	**43.96**	**1419.01**	**310.88**	**1388.40**	**17.72**

Table 6.16 Instances and proportions of objects of 同意 *(tóng yì)* and 赞同 *(zàn tóng)*

To summarize, the analysis has shown that the members of both English and Chinese candidate synonymous pairs differ in respect of the grammatical patterns and functions they favour, especially when the sub-categorizations of the subjects and objects are looked at. In other words, these near-synonyms are primed differently for colligations.

6.4.2.3. pragmatic association

6.4.2.3.1. negation

Again as with the English data, negation was the first feature I looked at with respect to the pragmatic

associations of the Chinese synonymous pair 同意 *(tóng yì)* and 赞同 *(zàn tóng)*. The analysis of the two words shows that 同意 *(tóng yì)* and 赞同 *(zàn tóng)* differ with respect to the strength of their co-occurrence with negation, with 24% and 16.8% of instances respectively occurring with some markers of negation. In addition, the two words also differ in distribution with respect to the use of different negative forms and negative items, such as 不 *(bù) (not),* 未必 *(wèi bì) (not)* and 没有人 *(méi yǒu rén) (nobody)*(see Table 6.17).

	Negation (total)	不 *(bù)* (not)	未必 *(wèi bì)* (not necessarily)	没有人 *(méi yǒu rén)*(nobody)
同意(*tóng yì*) 250 instances	60 (24%)	59 (98.3%)	1 (1.7%)	0
赞同(*zàn tóng*) 250 instances	42 (16.8%)	39 (92.9%)	2 (4.8%)	1 (2.4%)
LL	**205.67**	**263.51**	**24.13**	**172.06**

Table 6.17 instances and proportions of negation with *同意 (tóng yì) and 赞同 (zàn tóng)*

6.4.2.3.2. elicitation or confirmation of opinions

Note that there are two singular forms for the second person pronouns in Chinese including 你 *(nǐ)* and 您 *(nín, respect form)*. In the structure 你 *(nǐ)*/您 *(nín)*+ 同意 *(tóng yì)*, 7 out of 13 instances were used for eliciting opinions whereas 2 out of 4 were found in the pattern 你 *(nǐ)*/您 *(nín)* +赞同 *(zàn tóng)*. As before, the data here do not permit confident conclusions to be drawn, but point to a difference that might be investigated more fully with more data. To be brief, the two synonymous pairs both in English and Chinese have shown potential differences in negation and eliciting/confirming opinions, in other words, potential pragmatic associations.

6.5. Conclusions and limitations

The aim of this section has been to test whether Chinese synonyms whether Chinese near-synonyms are primed differently in terms of their collocations, colligations, semantic associations and pragmatic associations. The research questions are, as noted earlier, (1) Are members of English near-synonymous verb pairs or sets primed differently for collocation, semantic association, colligation and pragmatic association? (2) Is it also the same true of Chinese near-synonyms? And (3) If it works with Chinese, are there any similarities and differences between Chinese and English near-synonyms in terms of collocation, semantic association, colligation and pragmatic association?

The analysis of the English synonymous verb pair *AGREE* and *CONCUR* has shown that the two verbs share similarities in their collocations (for instance, with prepositions and pronouns) and semantic groups (e.g. adverbs), but differ in the strength of association with respect to different collocates within the same semantic set. When the Subjects and Objects of *AGREE* and *CONCUR* are looked at, the two

verbs appear to favour different grammatical patterns and functions; here there is less similarity. Likewise, the analysis of the two words also shows divergence when different pragmatic functions are concerned.

As with the English pair, the Chinese synonymous verbs 同意 *(tóng yì)* and 赞同 *(zàn tóng)* have shown similar behaviours. In association with prepositions and pronouns, they share similar collocates and semantic groups but differ in the distributions with regard to individual collocates in the same semantic set. The description of grammatical patterns in Chinese is very different from that of English; the analysis of Chinese data however presents us a similar picture with English data. The Chinese synonyms 同意 *(tóng yì)* and 赞同 *(zàn tóng)* also tend to occur with different distribution as regards various grammatical patterns and divergent grammatical functions. The small number of data in the analysis of two words in terms of pragmatic associations are insufficient to support the claim but provide a direction for further research.

One point worth mentioning here is that based on the log-likelihood statistics the Chinese synonyms 同意 *(tóng yì)* and 赞同 *(zàn tóng)* seem more divergent than what we might have expected of candidate synonyms while the English set *AGREE* and *CONCUR* seem to be more close to our expectations. A possible explanation may be that *AGREE* and *CONCUR* are more similar to each other than the Chinese pair 同意 *(tóng yì)* and 赞同 *(zàn tóng)*. However, we cannot exclude other possibilities. As Hanks (2002) observes, a word 'can have about as many senses as the lexicographer cares to perceive'. The categorisation of senses is therefore to some extent subjective. Another possibility concerns the part-of-speech tagging of the data. Due to the abundance and maturity of English research studies, the accuracy rate in English tagging is high and reliable. However, with the Chinese data, POS tagging is much more complex due to the fact that Chinese is not an inflectional language and thus forms are not of any help in deciding the part-of-speech. Because Chinese corpus studies are still in their infancy, the accuracy of both tagging and discrimination of word senses need to be improved. Thus we have to bear in mind that the statistics offered in this chapter cannot be absolutely reliable because the figures are based on decisions about part-of-speech allocation and the sense distinction that may not be fully reliable.

Despite the deficiencies in the statistics, the analysis of both English and Chinese near-synonymous verbs has supported the hypothesis that the claim concerning synonymy in lexical priming has universal applications, which has been demonstrated with the analysis of two different typological languages. In addition, as lexical priming is capable of measuring the strength of similarities between synonymous nouns in English and verbs (my analysis) both in English and Chinese, it seems appropriate to test the notion of synonymy in Chinese within the framework of lexical priming with a focus on the strength of similarity, which will be the focus of the next chapter.

Chapter 7 A Corpus-Driven Investigation into Collocational and Colligational Behaviours of Potentially Synonymous Items in Chinese

7.1. Introduction to the chapter

In the previous chapters the concept of synonymy was examined from a psycholinguistic perspective and using corpus analysis on English and, to a lesser extent, on Chinese. The first two research questions and the fifth one (partially) have been addressed, namely:

(1) How do people understand the notion of synonymy? Does synonymy have psychological reality? Do people share, or differ in, the meaning they ascribe to their sense of synonymy?

(2) If we find that synonymy has psychological reality, does the analysis of corpus data help to explain the findings obtained in my psychological experiment?

(5) Given Cruse (2002)'s claims that synonymy is scalar, do the categories used in Lexical Priming help us to measure the strength of synonymy between pairs or among a group of words in the two unrelated languages?

The results of the psycholinguistic experiment show that people have the concept of synonymy but they differ in the meaning they ascribe to the notion, and the corpus analysis of English candidate synonyms shows that synonymy is very complicated and that the term is too simplified, as in real use of the English language, the distinction between synonymy, hyponymy and even metaphor can be blurred, especially in various co-texts and contexts. We also have seen in the past two chapters that corpus linguistic methodology and, in particular, the key notions of Lexical Priming theory are capable of showing the strength of the synonymy between candidate synonyms.

We also saw that despite the close similarity of their dictionary definitions, 同意 (*tóng yì*) and 赞同 (*zàn tóng*) do not share many primings and it was argued that this shows they are not synonymous. This left however an unanswered question: does synonymy exist in Chinese in a form identifiable by the categories used to identify synonyms in English? Therefore it seems appropriate to look at more Chinese data within the framework of Lexical Priming. As shown in Chapter 5 synonymy is a complex language phenomenon which may not be easily defined and Lexical Priming may provide a possible way of locating synonymy. Having tested the applicability of Lexical Priming to the analysis of candidate Chinese synonyms, this chapter then seeks to extend the range of the research. The aim of the current chapter is to investigate whether it is possible to distinguish synonymy from co-hyponymy by looking at the collocational and colligational behaviours of a group of potential synonyms in Chinese. The results will be compared with those derived from the English data analysis in the previous chapter.

7.2. Methodology

7.2.1. choice of Chinese candidate words

As Chapter 5 looked at a group of potentially synonymous English words comprising *OUTCOME*,

IMPACT, AFTERMATH, UPSHOT, SEQUEL, EFFECT, END-PRODUCT, BY-PRODUCT, FRUIT, RESULT and *CONSEQUENCE*, ten potentially synonymous Chinese words were chosen for the analysis, namely 影响 (*yíng xiǎng*) (influence), 成果 (*chéng guǒ*) (achievement), 结果 (*jié guǒ*) (result), 效果 (*xiào guǒ*) (effect), 后果 (*hoù guǒ*) (consequence), 果实 (*guǒ shí*) (fruit), 结局 (*jié jú*) (end), 硕果 (*shuò guǒ*) (achievement), 恶果 (*è guǒ*) (evil fruit) and 苦果 (*kǔ guǒ*) (bitter fruit). As Xiao and McEnery (2006) has pointed out, English and Chinese have a different range of synonyms. To work with authentic data, the above words were chosen not based on the basis of English translation equivalence but on what Chinese people traditionally consider to be synonymous items in Chinese.

7.2.2. corpus and analysis tool

Again the Sketch Engine was used to analyse the data in the zhTenTen11 corpus. Using the same analytical tools and comparable corpora (both BNC and zhTenTen 11 are general corpora) helps us legitimately pair the results of analyses in the two corpora. However, it must be borne in mind that, as noted in the previous chapter, the Chinese zhTenTen 11 is a web-crawling corpus and it is difficult to trace the genre/text type of every concordance line in it, so conclusions about genre/text type must be drawn with caution.

As with the analysis of the English data, the analysis of the Chinese data follows the same sequence and uses the same measures and concepts, namely frequency (raw and standardised), collocation, semantic association and colligation.

7. 3. Results and analysis

7.3.1. Frequency

The first investigation concerns the relative frequency of the candidate synonyms. We note again that Chinese is a non-inflectional language, which means there is no variation of word form, so the notion of lemma is not applicable here. I used the word form entry and chose the noun form of each word. If we take 影响 (*yíng xiǎng*) (influence) as an example, the snapshot of the search entry looks as follows in Figure 7.1.

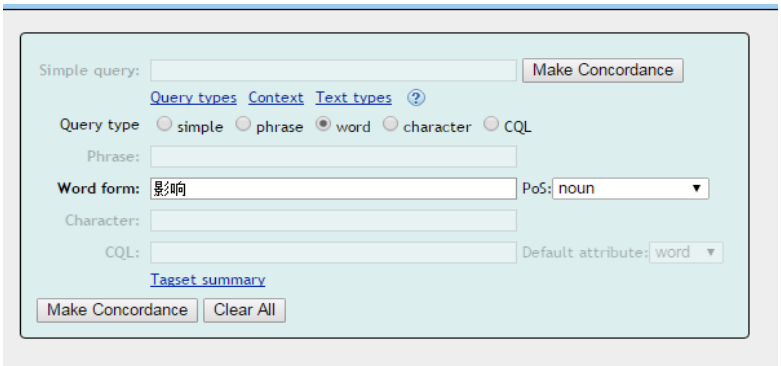


Figure 7.1 Snapshot of search entry in the Sketch Engine (take 影响 (*yíng xiǎng*)(influence) as an example)

In the zhTenTen11, there are 751,927 (356.9 per million) instances of 影响 (*yíng xiǎng*) (influence), 676,635 (321.2 per million) instances of 成果 (*chéng guǒ*) (achievement), 462,769 (219.7 per million) instances of 结果 (*jié guǒ*) (result), 433,393 (205.7 per million) instances of 效果 (*xiào guǒ*) (effect), 68,006 (32.3 per million) instances of 后果 (*hoù guǒ*) (consequence), 19,130 (9.1 per million) instances of 果实 (*guǒ shí*) (fruit), 14,158 (6.7 per million) instances of 结局 (*jié jú*) (end), 5,176 (2.5 per million) instances of 硕果 (*shuò guǒ*) (achievement), 3,136 (1.5 per million) instances of 恶果 (*è guǒ*) (evil fruit) and 1,999 (0.6 per million) instances of 苦果 (*kǔ guǒ*) (bitter fruit). The words are ranked in terms of standardised frequency of the words in zhTenTen11 (Table 7.1).

Rank	Words in query	Raw frequency	Standardised Frequency (per million)
1	**影响** (*yíng xiǎng*) (influence)	751.927	356.9
2	**成果** (*chéng guǒ*) (achievement)	676,635	321.2
3	**结果** (*jié guǒ*) (result)	462,769	219.7
4	**效果** (*xiào guǒ*) (effect)	433.393	205.7
5	**后果** (*hoù guǒ*) (consequence)	68,006	32.3
6	**果实** (*guǒ shí*) (fruit)	19,130	9.1
7	**结局** (*jié jú*) (end)	14,158	6.7
8	**硕果** (*shuò guǒ*) (achievement)	5,176	2.5
9	**恶果** (*è guǒ*) (evil fruit)	3,136	1.5
10	**苦果** (*kǔ guǒ*) (bitter fruit)	1,999	0.6

Table 7.1 Raw and standardised frequency of the lemmas in the zhTenTen11 corpus

Because Chinese is a non-inflectional language and zhTenTen11 is a web-crawling corpus, it is difficult to trace the genre/text type of every concordance line in it; therefore no analysis concerning different word forms or genre/text type was conducted with the Chinese data.

7.3.2. collocation

As with the English data, I used Word Sketch to retrieve the collocational behaviours of the potentially synonymous words in Chinese. The result does not seem to show various categories of collocates with different lexical categroies and functions as with English data (see Chapter 5.2.2). Word Sketch only elicited two categories of collocates for all the candidate words: modifiers and modifieds. The former category includes words which are functioning as modifiers of the query word and the latter consists of those which modify the word in query. The result may be due to the issues of grammatical categorisation

and automatic segmentation in Chinese.

Take 影响 (*yíng xiǎng*) (influence) as an example. Table 7.2 shows all the collocates functioning as modifiers of 影响 (*yíng xiǎng*) (influence) in the order of their collocational significance.

Collocates	Frequency	Significance
深远 (*shēn yuǎn*) (deep, profound)	6,868	9.85
因素 (*yīn sù*) (factor)	7,002	9.15
带来 (*dài lái*) (bring about)	4,647	8.94
危机 (*wēi jī*) (crisis)	6,610	8.90
大 (*dà*) (big)	11,441	8.90
产生 (*chǎn shēng*) (produce)	3,978	8.73
造成 (*zào chéng*) (cause)	3,406	8.65
环境 (*huán jìng*) (environment)	18,576	8.37
积极 (*jī jí*) (active, positive)	3,006	8.34
深刻 (*shēn kè*) (profound, incisive)	2,433	8.13
广泛 (*guǎng fàn*) (wide, extensive)	2,419	8.12
重要 (*zhòng yào*) (important)	3,877	7.96
巨大 (*jù dà*) (huge)	2,139	7.77
不利 (*bú lì*) (unfavourable)	1,495	7.70
重大 (*zhòng dà*) (major)	1,310	7.42
天气 (*tiān qì*) (weather)	1,521	7.33
潜移默化 (*qiǎn yí mò huà*) (silent and unconscious)	1,140	7.33
社会 (*shè huì*) (society)	13,773	7.24
受到 (*shòu dào*) (receive)	900	6.96
严重 (*yán zhòng*) (serious, critical)	1,076	6.95
恶劣 (*è luè*) (adverse)	863	6.90
冷空气 (*lěng kōng qì*) (cold weather)	819	6.85
坏 (*huài*) (bad)	710	6.60
地震 (*dì zhēn*) (earthquake)	1,001	6.54

Table 7.2 Collocates as modifiers of 影响 (*yíng xiǎng*) (influence)

A glance at the table shows that these modifiers can be further subcategorised according to their lexical category. After manually sorting the modifier collocates of 影响 (*yíng xiǎng*)(influence), three subcategories were identified. The first group is composed of words functioning as adjectives which are used to modify the query word. These words are 深远 (*shēn yuǎn*)(deep, profound), 积极 (*jī jí*)(active, positive), 广泛 (*guǎng fàn*)(wide, extensive), 大 (*dà*)(big), 深刻 (*shēn kè*)(profound, incisive), 重要 (*zhòng yào*)(important), 巨大 (*jù dà*)(huge), 不利 (*bú lì*)(unfavourable), 重大 (*zhòng dà*)(major), 严重 (*yán zhòng*)(serious, critical), 恶劣 (*è luè*)(adverse), 坏 (*huài*)(bad) and 潜移默化 (*qiǎn yí mò huà*)(silent and unconscious). Examples are given below.

Example 7.1:

这 对 中国 的 建筑业 也 必 将 产生 ***深远*** 的 **影响** 。
Zhè duì zhōng guó de jiàn zhù yè yě bì jiāng chǎn shēng ***shēn yuǎn*** de **yǐng xiǎng**.
This towards China PAR construction industry also will produce deep PAR influence.
This will also have a deep influence on China's construction industry.

Example 7.2:

这样 的 活动 是否 对 你 的 学习 与 生活 产生 了 ***积极*** 的 **影响**?
Zhè yang de huó dòng shì fǒu duì nǐ de xué xí yǔ shēng huó chǎn shēng le ***jī jí*** de **yǐng xiǎng**?
This PAR activity yes no towards you PAR study and life produce PAR positive PAR influence.
Does this kind of activity have a positive influence on your study and life?

The second group consists of words functioning as verbs such as 产生 (*chǎn shēng*)(produce), 带来 (*dài lái*)(bring about), 造成 (*zào chéng*)(cause) and 受到 (*shòu dào*)(receive). These verbs, together with the word 的 (*de*)(functional word), are modifiers of the noun. Examples are:

Example 7.3:

在 不同 的 地方， 气候 变化 ***带来*** 的 物理 **影响** 会 有 所 差异。
Zài bù tóng de dì fāng, qì hou biàn huà ***dài lái*** de wù lǐ **yǐng xiǎng** huì yǒu suǒ chā yì.
At different PAR place, climate change bring PAR physical influence will have difference.
There will be differences in the physical influence brought about by climate change at different places.

Example 7.4:

上述 因素 对 粮食 需求 ***产生*** 的 **影响** 相当 惊人 。
Shàng shù yīn sù duì liáng shi xū qiú ***chǎn sheng*** de **yǐng xiǎng** xiāng dāng jīng rén.
Above mentioned factor towards food demand produce PAR influence considerably surprising.
The influence on food demand produced by the above mentioned factors is considerably surprising.

Thirdly, words functioning as nouns include 环境 (*huán jìng*)(environment), 天气 (*tiān qì*)(weather), 因素 (*yīn sù*)(factor), 危机 (*wēi jī*)(crisis), 社会 (*shè huì*)(society), 冷空气 (*lěng kōng qì*)(cold

weather) and 地震 (*dì zhēn*)(earthquake). Examples are:

Example 7.5:

受 寒冷 ***天气*** **影响** ， 物价 短期 无 下降 可能 。
Shòu hán lěng ***tiān qì*** **yǐng xiǎng**, wù jià duǎn qī wú xià jiàng kě néng.
Suffer cold weather influence, price short time no decrease possibility.
Influenced by cold weather, there is no possibility of price decrease for now. (in the short term?)

Example 7.6:

受 金融 ***危机*** **影响** ， 该 公司 于 去年 年底 被 迫 停产 。
Shòu jīn róng ***wēi jī*** **yǐng xiǎng**, gāi gōng sī yǔ qù nián bèi pò tíng chǎn.
Suffer financial ***crisis*** **influence**, this company at last year be forced stop production.
Influenced by financial crisis, this company was forced to stop production last year.

Note that the English translation may suggest 影响 (*yíng xiǎng*) (influence) is used as a past particle of the verb. However, since there is no inflectional variation in Chinese, it is very difficult to decide the lexical category from the word form. For example, 学习 (*xué xí*) (study) can be both noun and verb. We have to look at the sentence in which the word is used to decide the lexical category and its function. In the following examples, 学习 (*xué xí*) is used as verb in sentence 7.7 and as a noun in sentence 7.8.

Example 7.7:

我们 一定 要 认真 ***学习*** 、 深刻 领会 。
Wǒ men yí dìng yào rèn zhēn ***xué xí***, shēn kè tǐ huì.
We must seriously study, profound comprehend.
We must study it carefully and comprehend it profoundly.

Example 7.8:

我们 应 不断 改善 教师 的 工作 、 ***学习*** 和 生活 条件。
Wǒ men yīng bú duàn gǎi shàn jiào shi de gong zuò, ***xué xí*** hé shēng huó tiáo jiàn.
We should continually improve teacher PAR work, study and live condition.
We should continually improve the teacher's working, studying and living conditions.

7.3.2.1. adjective collocates

Following the procedure used in Chapter 5, I classified the modifier collocates of all the candidate synonyms. Table 7.3 lists the adjective collocates which are functioning as all the modifiers of the words in query. The frequency and significance of the collocational association between the candidate words and their adjective collocates are also provided in Table 7.3.

Surprisingly, for 结果 (*jié guǒ*) (result) there are no adjective collocates functioning as modifiers, but many noun collocates (discussed in a later section). In addition, no adjective collocates for 苦果 (*kǔ guǒ*) (bitter fruit) have been found either. One might argue that as the character 苦 (*kǔ*) means *bitter*,

which can be considered to be the modifier of 果 (*guǒ*)(fruit), no more modifiers are needed. However, it may also be argued that although in 恶果 (*è guǒ*)(evil fruit), the character 恶 (*è*)(evil) could also be considered to be a modifier of 果 (*guǒ*)(fruit), adjective collocates such as 严重 (*yán zhòng*)(serious, critical) and 可怕 (*kě pà*)(terrible) have still been found in the data.

Rank	Words in query	Collocates (functioning adjective as modifiers)	Frequency (Significance)
1	影响 (*yíng xiǎng*) (influence)	深远 (*shēn yuǎn*) (deep, profound)	6,868 (9.85)
		积极 (*jī jí*) (active, positive)	3,006 (8.33)
		广泛 (*guǎng fàn*) (wide, extensive)	2,419 (8.11)
		大 (*dà*) (big)	11,441 (8.90)
		深刻 (*shēn kè*) (profound, incisive)	2,433 (8.13)
		重要 (*zhòng yào*) (important)	3,877 (7.95)
		巨大 (*jù dà*) (huge)	2,139 (7.77)
		不利 (*bú lì*) (unfavourable)	1,495 (7.70)
		重大 (*zhòng dà*) (major)	1,310 (7.42)
		严重 (*yán zhòng*) (serious, critical)	1,076 (6.95)
		恶劣 (*è luè*) (adverse)	863 (6.90)
		坏 (*huài*) (bad)	710 (6.60)
		潜移默化 (*qiǎn yí mò huà*) (silent and unconscious)	1,140 (7.32)
2	成果 (*chéng guǒ*) (achievement)	丰硕 (*fēng shuò*) (rich, plentiful)	5,949 (8.69)
		阶段性 (*jiē duàn xìng*) (of stage, of phase)	4,324 (8.19)
		优秀 (*yōu xiù*) (excellent)	2,666 (7.29)
		新 (*xīn*) (new)	2,828 (6.59)
		好 (*hǎo*) (good)	1,877 (6.03)
3	结果 (jié guǒ) (result)	/	/
4	效果 (*xiào guǒ*) (effect)	好 (*hǎo*) (good),	21,727 (9.89)
		良好 (*liáng hǎo*) (fine)	15,062 (9.74)
		明显 (*míng xiǎn*) (obvious)	2,241 (7.80)
		实实在在 (*shí shí zài zài*) (substantial)	1,623 (7.55)
		满意 (*mǎn yì*) (satisfactory)	1,908 (7.54)
		意想不到 (*yì xiǎng bú dào*) (unexpected)	1,172 (7.12)

		显著 (*xiǎn zhù*) (notable, outstanding)	1,183 (7.03)
		理想 (*lí xiǎng*) (ideal)	1,252 (6.82)
		不错 (*bú cuò*) (not bad)	896 (6.65)
		佳 (*jiā*) (good, fine)	949 (6.58)
5	后果 (*hoù guǒ*) (consequence)	严重 (*yán zhòng*) (serious, critical)	1,927 (9.58)
		不利 (*bú lì*) (unfavourable)	285 (8.15)
		可怕 (*kě pà*) (terrible)	282 (8.10)
		不良 (*bù liáng*) (not good)	156 (7.31)
		直接 (*zhí jiē*) (direct)	124 (6.87)
		意想不到 (*yì xiǎng bú dào*) (unexpected)	79 (6.33)
6	果实 (*guǒ shí*) (fruit)	丰硕 (*fēng shuò*) (rich, plentiful)	623 (10.48)
		胜利 (*sheng lì*) (victory, successful)	549 (8.71)
		丰收 (*fēng shōu*) (harvest)	95 (8.20)
		甜美 (*tián měi*) (sweet and nice)	76 (8.17)
		累累 (*lěi lěi*)(clusters of)	36 (7.44)
		成熟 (*chéng shú*) (ripe)	235 (7.15)
		红红 (*hóng hóng*) (red)	16 (6.16)
		甜蜜 (*tián mì*) (sweet)	21 (6.00)
		沉甸甸 (*chén diān dian*) (heavy)	14 (6.00)
		恶魔 (*è mó*) (evil monster)	14 (5.97)
7	结局 (*jié jú*) (end)	悲惨 (*bēi cǎn*) (miserable)	187 (9.55)
		圆满 (*yuán mǎn*) (satisfactory)	162 (8.41)
		完满 (*wán mǎn*) (satisfactory)	36 (7.76)
		完美 (*wán měi*) (perfect)	280 (7.50)
		出人意料 (*chū rén yì liào*) (unexpected)	23 (7.17)
		可悲 (*hě bēi*) (pitiable)	21 (7.04)
		必然 (*bì rán*) (inevitable)	29 (6.41)
		悲凉 (*bēi liáng*) (sad)	14 (6.41)
		糟糕 (*zāo gāo*) (bad)	16 (6.16)
		坏 (*huài*) (bad)	31 (6.11)

		凄惨 (*qī cǎn*) (miserable)	10 (6.06)
		无奈 (*wú nài*) (have to choice)	21 (5.99)
8	**硕果** (*shuò guǒ*) (achievement)	**累累** (*lěi lěi*) (clusters of)	98 (10.18)
		丰收 (*fēng shōu*) (harvest)	47 (7.88)
		喜人 (*xǐ rén*) (gratifying)	7 (6.11)
		沉甸甸 (*chén diān dian*) (heavy)	3 (4.96)
		丰厚 (*fēng hòu*) (rich)	10 (4.78)
		丰盛 (*fēng shèng*) (rich, bumper)	4 (4.47)
9	**恶果** (*è guǒ*) (evil fruit)	**严重** (*yán zhòng*) (serious, critical)	41 (4.55)
		可怕 (*kě pà*) (terrible)	4 (3.93)
10	**苦果** (*kǔ guǒ*) (bitter fruit)	/	/

Table 7.3 Collocates of the candidate synonyms functioning as adjective modifiers

Based on the adjective collocates which are used to modify the words in query, 硕果 (*shuò guǒ*) (achievement), 成果 (*chéng guǒ*)(achievement) and 果实 (*guǒ shí*)(fruit) typically express positive meaning while 结局 (*jié jú*) (end), 后果(*hoù guǒ*)(consequence), 恶果 (*è guǒ*)(evil fruit) and 影响 (*yíng xiǎng*) (influence) have negative connotations. 效果 (*xiào guǒ*) (effect) is neutral with both a positive collocate 好 (*hǎo*) (good) and neutral 意想不到 (*yì xiǎng bú dào*) (unexpected).

As with English word *fruit*, when we look at the adjective collocates of the word 果实 (*guǒ shí*)(fruit), namely 丰硕 (*fēng shuò*)(rich, plentiful), 胜利 (*sheng lì*)(victory, successful), 丰收 (*fēng shōu*)(harvest), 甜美 (*tián měi*)(sweet and nice), 累累 (*lěi lěi*)(clusters of), 成熟 (*chéng shú*)(ripe), 红红 (*hóng hóng*) (red), 甜蜜 (*tián mì*)(sweet), 沉甸甸 (*chén diān dian*)(heavy) and 恶魔 (*è mó*)(evil monster), it would appear that the word 果实 (*guǒ shí*)(fruit) has both a literal sense and a metaphorical sense. The adjective collocates 胜利 (*sheng lì*)(victory, successful) and 恶魔 (*è mó*)(evil monster) seem to associate with the metaphorical sense of 果实 (*guǒ shí*)(fruit) while the others may be related to the literal sense, and 丰硕 (*fēng shuò*)(rich, plentiful) is possible to associate with both senses.

Table 7.4 lists all the shared collocates as adjective modifiers. It can be seen that the candidate words are sharing adjective collocates in an intertwined way. For example, 果实 (*guǒ shí*)(fruit) shares the collocate 丰硕 (*fēng shuò*)(rich, plentiful) with 成果 (*chéng guǒ*)(achievement), and shares the collocates 丰收 (*fēng shōu*)(harvestful), 累累 (*lěi lei*)(clusters of) and 沉甸甸 (*chén diān dian*)(heavy) with 硕果 (*shuò guǒ*)(achievement). The sharing of these collocates may suggest a

closeness of meanings/senses among these candidate synonyms. However it does not tell us whether any two of these words are synonymous and or whether others are not. As has been discussed in Chapter 5, it is possible that polysemous senses of the words may compromise attempts to measure the strength of similarities among the candidate words.

Collocates	shared by …
丰硕 (*fēng shuò*) (rich, plentiful)	果实 (*guǒ shí*) (fruit), 成果 (*chéng guǒ*) (achievement)
丰收 (*fēng shōu*) (harvestful) ,	果实 (*guǒ shí*) (fruit), 硕果 (*shuò guǒ*) (achievement)
累累 (*lěi lei*) (clusters of)	果实 (*guǒ shí*) (fruit), 硕果 (*shuò guǒ*) (achievement)
沉甸甸 (*chén diān dian*) (heavy)	果实 (*guǒ shí*) (fruit), 硕果 (*shuò guǒ*) (achievement)
好 (hǎo) (good)	成果 (*chéng guǒ*) (achievement), 效果 (*xiào guǒ*) (effect)
坏 (*huài*) (bad)	影响 (*yíng xiǎng*) (influence), 结局 (*jié jú*) (end)
不利 (*bú lì*) (unfavourable)	影响 (*yíng xiǎng*) (influence), 后果 (*hoù guǒ*) (consequence)
可怕 (*kě pà*) (terrible)	后果 (*hoù guǒ*)(consequence) , 恶果 (*è guǒ*)(evil fruit)
意想不到 (*yì xiǎng bú dào*) (unexpected)	效果 (*xiào guǒ*) (effect), 后果 (*hoù guǒ*) (consequence)
严重 (*yán zhòng*) (serious, critical)	影响 (*yíng xiǎng*) (influence), 后果 (*hoù guǒ*) (consequence), 恶果 (*è guǒ*) (evil fruit)

Table 7.4 Shared collocates of the words in query functioning as adjective modifiers

7.3.2.2. verb collocates

Next, the verb collocates of each word in query are listed in Table 7.5. Again for 结果 (*jié guǒ*) (result) there are no verb collocates functioning as modifiers. However, 苦果 (*kǔ guǒ*)(bitter fruit) has elicited the most verb collocates, namely 种下 (*zhòng xià*)(sow), 咽 (*yàn*)(swallow), 酿下 (*niàng xià*)(brew), 酿成 (*niàng chéng*)(breed), 自酿 (*zì niàng*)(brew), 酿就 (*niàng jiù*)(brew), 下咽 (*xia yàn*)(swallow), 酿出 (*niàng chū*)(brew), 结下 (*jié xià*)(bear), 酿造 (*niang zào*)(brew), 种出 (*zhòng chū*)(sow), 砸 (*zá*)(smash), 埋下 (*mái xià*)(bury), 结出 (*jié chū*)(bear) and 吞咽 (*tūn*

yàn)(swallow).

Rank	Words in query	Collocates (verbs)	Frequency (Significance)
1	影响 (*yíng xiǎng*) (influence)	产生 (*chǎn shēng*) (produce)	3,978 (8.72)
		带来 (*dài lái*) (bring about)	4,647 (8.93)
		造成 (*zào chéng*) (cause)	3,406 (8.65)
		受到 (*shòu dào*) (receive)	900 (6.96)
2	成果 (*chéng guǒ*) (achievement)	取得 (*qǔ dé*) (gain, obtain)	11,733 (9.46)
3	结果 (*jié guǒ*)(result)	/	
4	效果 (*xiào guǒ*) (effect)	预期 (*yù qī*) (expect)	4,090 (8.74)
		达到 (*dá dào*) (reach, attain)	875 (6.68)
5	后果 (*hoù guǒ*) (consequence)	导致 (*dǎo zhì*) (lead to)	531 (8.65)
		产生 (*chǎn shēng*) (produce)	1,093 (8.41)
		引起 (*yǐn qǐ*) (give rise to)	312 (7.61)
		承担 (*chéng dān*) (undertake)	159 (6.66)
6	果实 (*guǒ shî*) (fruit)	结出 (*jié chū*) (bear)	93 (8.77)
		结 (*jié*) (bear)	35 (6.24)
7	结局 (*jié jú*) (end)	避免 (*bì miǎn*) (avoid)	20 (6.25)
		看到 (*kàn dào*) (see)	39 (6.18)
8	硕果 (*shuò guǒ*) (achievement)	结出 (*jié chū*) (bear)	171 (10.88)
		结 (*jié*) (bear)	125 (8.52)
		结下 (*jié xià*) (bear)	16 (7.74)
		结成 (*jié chéng*) (bear)	5 (5.97)
		收获 (*shōu huò*) (harvest)	57 (5.81)
		换来 (*huàn lái*) (exchange)	4 (5.41)
		存 (*cún*) (store)	6 (4.90)
9	恶果 (*è guǒ*) (evil fruit)	种下 (*zhòng xià*) (sow)	27 (8.96)
		招引 (*zhāo yǐn*) (induce)	19 (8.51)
		酿成 (*niàng chéng*) (breed)	17 (8.17)
		结出 (*jié chū*) (bear)	11 (7.42)
		造成 (*zào chéng*) (cause)	237 (7.28)

		挽回 (*wǎn huí*) (retrieve)	10 (6.93)
		导致 (*dǎo zhì*) (lead to)	58 (6.70)
		埋下 (*mái xià*) (bury)	3 (6.11)
		带来 (*dài lái*) (bring about)	146 (5.85)
		留下 (*liú xià*) (leave)	6 (4.06)
		引起 (*yǐn qǐ*) (give rise to)	13 (3.98)
		引发 (*yǐn fā*) (trigger)	8 (3.96)
10	苦果 (*kǔ guǒ*) (bitter fruit)	种下 (*zhòng xià*) (sow)	29 (10.01)
		咽 (*yàn*) (swallow)	9 (9.21)
		酿下 (*niàng xià*) (brew)	6 (8.88)
		酿成 (*niàng chéng*) (breed)	15 (8.85)
		自酿 (*zì niàng*) (brew)	8 (8.76)
		酿就 (*niàng jiù*) (brew)	3 (7.86)
		下咽 (*xià yàn*) (swallow)	2 (7.11)
		酿出 (*niàng chū*) (brew)	2 (7.07)
		结下 (*jié xià*) (bear)	3 (7.05)
		酿造 (*niang zào*) (brew)	9 (6.97)
		种出 (*zhòng chū*) (grow)	2 (6.96)
		砸 (*zá*) (smash)	2 (6.90)
		埋下 (*mái xià*) (bury)	2 (6.90)
		结出 (*jié chū*) (bear)	3 (6.30)
		吞咽 (*tūn yàn*) (swallow)	2 (6.11)

Table 7.5 Collocates of the candidate synonyms functioning as verb modifiers

As noted several times, it is not easy to distinguish lexical categories in Chinese. The verbs classified here usually establish a verb-object relation with the word in query. Consider the following example:

Example 7.9:

此前　高铁 建设　对 公路　交通　***造成*** 的　*影响*　被 高 估。
Cǐ qián gāo tiě jiàn shè duì gōng lù jiāo tōng ***zào chéng*** de *yíng xiǎng* bèi gāo gū.
This before high speed rail construction towards road traffic caused PAR influence PAR over-estimated.
The influence on road traffic caused by high speed rail construction has been over-estimated before.

In this example, the word 影响 (*yíng xiǎng*) (influence) and the verb 造成 (*zào chéng*)(cause) are in a verb-object relation. However look at the following example:

Example 7.10:
该 书 是 众多 专家 学者 数 年 埋头 ***耕耘*** 的 *硕果*。
Gāi shū shū zhòng duō zhuān jiā xué zhě shù nián mái tóu ***gēng yún*** de *shuò guǒ*.
This book is numerous expert scholar many year bury head cultivate PAR achievement.
This book is the achievement after many years of cultivation of numerous experts and scholars.

Here it would appear that 硕果 (*shuò guǒ*) (achievement) is not the object of 耕耘 (*gēng yún*) (cultivate, as a verb) as it is not the object that receives the direct action of the verb, and therefore 耕耘 (*gēng yún*) is categorised as a noun collocate (see Table 7.7).

In Table 7.6, all the shared verb collocates are listed. The shared verb collocates seem to suggest the closeness of senses among the words. For example, 影响 (*yíng xiǎng*) (influence) share collocates 带来 (*dài lái*) (bring about), 产生 (*chǎn shēng*) (produce) and 造成 (*zào chéng*) (cause) with 恶果 (*è guǒ*)(evil fruit) and 后果 (*hoù guǒ*)(consequence). Again this suggests the closeness in sense of the candidate words; it however does not show whether the words are synonymous or not because this could be treated as co-hyponyms.

Collocate	Shared by
带来 (*dài lái*) (bring about)	影响 (*yíng xiǎng*) (influence), 恶果 (*è guǒ*) (evil fruit)
产生 (*chǎn shēng*) (produce)	影响 (*yíng xiǎng*) (influence), 后果 (*hoù guǒ*) (consequence)
造成 (*zào chéng*) (cause)	影响 (*yíng xiǎng*) (influence), 恶果 (*è guǒ*) (evil fruit)
引起 (*yǐn qǐ*) (give rise to)	后果 (*hoù guǒ*) (consequence) 恶果 (*è guǒ*) (evil fruit)
结出 (*jié chū*) (bear)	果实 (*guǒ shí*) (fruit) 硕果 (*shuò guǒ*) (achievement) 苦果 (*kǔ guǒ*) (bitter fruit)
结 (*jié*) (bear)	果实 (*guǒ shí*) (fruit) 硕果 (*shuò guǒ*) (achievement)
结下 (*jié xià*) (bear)	硕果 (*shuò guǒ*) (achievement) 苦果 (*kǔ guǒ*)(bitter fruit)
种下 (*zhòng xià*) (sow)	恶果 (*è guǒ*)(evil fruit) , 苦果 (*kǔ guǒ*)(bitter fruit)

埋下 (*mái xià*) (bury)	**恶果** (*è guǒ*) (evil fruit)， **苦果** (*kǔ guǒ*)(bitter fruit)
酿成 (*niàng chéng*) (breed)	**恶果** (*è guǒ*) (evil fruit)， **苦果** (*kǔ guǒ*)(bitter fruit)

Table 7.6 Shared collocates of the candidate synonyms, functioning as verb modifiers

7.3.2.3. noun collocates

We now move on to the noun collocates as modifiers. Table 7.7 lists all the noun collocates of the words in query. As before, there are difficulties in the grammatical categorisation. Consider the following example.

Example 7.11:

南方　有些　地区 因 受到　***天气***　*影响*　票房　受到　不 小 损失。

Nán fāng yǒu xiē dì qū yīn shòu dào ***tiān qì*** *yíng xiǎng* piào fáng shòu dào bù xiǎo sǔn shī。

South some area because PASSIVE weather influence ticket sell PASSIVE not small influence.

The ticket selling in some area in the South suffered a big loss because of the influence of the weather.

In this example, 天气 (*tiān qì*)(weather) and 影响 (*yíng xiǎng*)(influence) seem to have a subject-predicate (or entity-action) relation, in which 影响 (*yíng xiǎng*)(influence) appears to be the predicate. However, the noun collocates of other words in query do not share this feature.

Example 7.12:

调查　*结果*　显示 ， 出租车 司机 普遍 认为 费 用 过 高。

diào chá *jié guǒ* xiǎn shì, chū zū chē sī jī pǔ biàn rèn wéi fèi yòng guò gāo.

Investigation result show, taxi driver generally think cost over high.

The result of the investigation shows that most of the taxi drivers think the cost is too high.

调查 (*diào chá*)(investigation) in the above example is considered to be a noun collocate of 结果 (*jié guǒ*)(result) and seems to have similar grammatical function as that in English.

Table 7.7 lists all the noun collocates of the words in query. As has mentioned before, 结果 (*jié guǒ*) (result) does not have any adjective or verb collocates, but it has a long list of noun modifier collocates. Interestingly all the noun collocates of 果实 (*guǒ shí*)(fruit) are actually the names of different fruits, such as 蓝莓 (*lán méi*) (blueberry), 草莓 (*cǎo méi*) (strawberry), 荔枝 (*lì zhī*) (lychee) and 猕猴桃 (*mí hóu táo*) (kiwi). This again is clear evidence of the literal sense of 果实 (*guǒ shí*)(fruit).

Rank	Lemma	Collocates (n. as modifiers)	Frequency (Signifcance)
1	**影响**	**环境** (*huán jìng*) (environment)	18576 (8.36)

	(*yíng xiǎng*) (influence)	天气 (*tiān qì*) (weather)	1521 (7.33)
		因素 (*yīn sù*) (factor)	7002 (9.15)
		危机 (*wēi jī*) (crisis)	6610 (8.90)
		社会 (*shè huì*) (society)	13773 (7.24)
		冷空气 (*lěng kōng qì*) (cold weather)	819 (6.85)
		地震 (*dì zhēn*) (earthquake)	1001 (6.54)
2	成果 (*chéng guǒ*) (achievement)	科研 (*kē yán*) (scientific research)	52698 (10.50)
		科技 (*kē jì*) (tecknology) (technology)	63354 (9.98)
		研究 (*yán jiū*) (research)	78871 (9.78)
		教学 (*jiào xué*) (teaching)	22264 (8.30)
		创新 (*chuàng xīn*) (innovation)	13728 (8.11)
		调研 (*diào yán*) (investigation)	4986 (7.92)
		学术 (*xué shù*) (academic)	7435 (7.83)
		技术 (*jì shù*) (skill)	15173 (7.11)
		测绘 (*cè huì*) (survey and draw)	2073 (7.07)
		理论 (*lǐ lùn*) (theory)	4504 (6.85
		劳动 (*láo dòng*) (labour)	2700 (6.70)
		课题 (*kè tí*) (research project)	2009 (6.60)
		学习 (*xué xí*) (study)	3204 (6.49)
		文明 (*wén míng*) (civilisation)	2432 (6.40)
		改革 (*gǎi gé*) (reform)	3687 (6.35)
		实践 (*shí jiàn*) (practice)	2980 (6.33)
		发展 (*fā zhǎn*) (development)	9244 (6.27)
		研发 (*yán fā*) (research and development)	1372 (6.08)
3	结果 (*jié guǒ*) (result)	调查 (*diào chá*) (investigation)	16139 (9.56)
		考核 (*kǎo hé*) (examination)	12490 (9.23)
		评估 (*píng gū*) (estimation)	6738 (8.60)
		搜索 (*sōu suǒ*) (search)	3655 (8.50)
		评选 (*píng xuǎn*) (choose through public appraisal)	4121 (8.41)
		检测 (*jiǎn cè*) (check)	4633 (8.30)

		实验 (*shí yàn*) (experiment)	6640 (8.29)
		评价 (*píng jià*) (evaluation)	5654 (8.22)
		测试 (*cè shì*) (test)	4154 (8.19)
		评审 (*píng shěn*) (evaluation and examination)	3831 (8.13)
		检查 (*jiǎn chá*) (check)	6775 (8.11)
		审计 (*shěn jì*) (auditing)	4911 (8.08)
		测评 (*cè píng*) (test and evaluation)	2753 (8.05)
		努力 (*nǔ lì*) (effort)	2875 (8.04)
		处理 (*chù lǐ*) (management)	4922 (7.91)
		分析 (*fēn xi*) (analysis)	4452 (7.89)
		考评 (*kǎo píng*) (test and assessment)	2426 (7.83)
		评议 (*píng yì*) (evaluation and discussion)	2635 (7.78)
		试验 (*shì yàn*) (experiment)	2726 (7.78)
		计算 (*jì suàn*) (calculation)	2191 (7.70)
		监测 (*jiān cè*) (monitoring and survey)	3019 (7.67)
		统计 (*tǒng jì*) (statistics)	3514 (7.64)
		选举 (*xuǎn jǔ*) (election)	2238 (7.63)
		检验 (*jiǎn yàn*) (check and test)	2180 (7.45)
		判决 (*pàn jué*) (court judgement)	1586 (7.35)
4	效果 (*xiào guǒ*) (effect)	治疗 (*zhì liáo*) (treatment)	4718 (8.70)
		教学 (*jiào xué*) (teaching)	17965 (8.11)
		学习 (*xué xí*) (studying)	6481 (7.77)
		宣传 (*xuān chuán*) (propaganda)	4247 (7.23)
		实施 (*shí shī*) (put into practice)	2474 (7.16)
		使用 (*shǐ yòng*) (adoption)	2096 (6.99)
		节能 (*jié néng*) (saving energy)	1656 (6.94)
		传播 (*chuán bō*) (propaganda)	1438 (6.90)
		防治 (*fang zhì*) (prevention and treatment)	1448 (6.77)
		培训 (*péi xùn*) (training)	3846 (6.70)
5	后果	空袭 (*kōng xí*) (air raid)	132 (7.26)

	(*hoù guǒ*) (consequence)	法律 (*fǎ lǜ*) (law)	3084 (6.90)
		死亡 (*sǐ wáng*) (death)	134 (6.59)
6	果实 (*guǒ shí*) (fruit)	番茄 (*fān qié*) (tomato)	59 (7.37)
		蓝莓 (*lán méi*) (blueberry)	35 (7.33)
		鸭梨 (*yā lí*) (pear)	25 (6.90)
		沙棘 (*shā jí*) (sea-buckthorn)	31 (6.88)
		草莓 (*cǎo méi*) (strawberry)	42 (6.71)
		枇杷 (*pí bā*) (loquat)	24 (6.66)
		柑桔 (*gān jú*) (tangerine)，	25 (6.44)
		荔枝 (*lì zhī*) (lychee)，	23 (6.36)
		柑橘 (*gān jú*) (orange)	24 (6.09)
		杨梅 (*yang méi*) (bayberry)，	16 (6.03)
		猕猴桃 (*mí hóu táo*) (kiwi)	15 (5.89)
7	结局 (*jié jú*) (end)	团圆 (*tuán yuan*) (union)	99 (9.11)
		妊娠 (*jié jú*) (pregancy (pregnancy)	58 (7.45)
8	硕果 (*shuò guǒ*) (achievement)	丰收 (*fēng shōu*) (harvestful)	47 (7.88)
		耕耘 (*gēng yún*) (cultivation)	6 (6.00)
		金秋 (*jīn qīu*) (gold autumn)	20 (5.89)
		秋天 (*qīu tiān*) (autumn)	7 (5.27)
9	恶果 (*è guǒ*) (evil fruit)	吸毒 (*xī dú*) (taking drugs)	3 (5.76)
		泛滥 (*fàn làn*) (overflow)	4 (5.39)
		枯竭 (*kū jié*) (exhaustion)	3 (4.60)
		自由化 (*zì yóu huà*) (liberalization)	3 (4.47)
10	苦果(*kǔ guǒ*) (bitter fruit)	失利 (*shī lì*) (setback)	18 (8.69)
		蚀本 (*shí běn*) (losing one’s capital)	2 (7.23)
		惨败 (*cǎn bài*) (crushing defeat)	2 (6.85)
		落榜 (*luò bǎng*) (fail a competitive examination for a job or school admission)	2 (6.22)
		失败 (*shī bài*) (failure)	33 (6.19)
		后悔 (*hòu huǐ*) (regret)	2 (6.02)
		失恋 (*shī liàn*) (break-up)	2 (6.00)

		私有化 (*sī yǒu huà*) (privatization)	2 (5.87)
		枯竭 (*kū jié*) (exhaustion)	2 (4.36)

Table 7.7 Collocates of the candidate synonyms, functioning as noun modifiers

From Table 7.7, it seems that the candidate words do not share many noun collocates functioning as modifiers, except that 成果 (*chéng guǒ*)(achievement) and 效果 (*xiào guǒ*)(effect) share 教学 (*jiào xué*) (teaching) and 学习 (*xué xí*) (studying).

This section has looked at modifier collocates of the words in query; the corpus analysis seems to suggest these candidate synonyms share adjective and verb collocates but not noun collocates. As the words in query share collocates in an intertwined way, we could see the closeness of meaning/sense among these words. However, we could only say, for example, 后果 (*hoù guǒ*)(consequence) and 恶果 (*è guǒ*)(evil fruit) are more synonymous than 后果 (hoù guǒ)(consequence) and 效果 (xiào guǒ)(effect), as the former pair shares more collocates than the latter, but we could not say two words are synonyms while others are not. In addition, as with English word *fruit*, 果实 (*guǒ shí*)(fruit) is polysemous and fewer shared collocates with other candidate words may be due to the distortion of its polysemous senses. Now the corpus analysis seems to suggest that 结果 (*jié guǒ*)(result) is not a candidate synonym with any of the others. It may be argued that 结果 (*jié guǒ*)(result) is the superordinate of 成果 (*chéng guǒ*) (achievement), 后果 (*hoù guǒ*) (consequence), 硕果 (*shuò guǒ*) (achievement), 恶果 (*è guǒ*) (evil fruit) and 苦果 (*kǔ guǒ*) (bitter fruit). It would be interesting to conduct a psycholinguistic experiment with Chinese participants to see whether they would provide 结果 (*jié guǒ*)(result) as the candidate synonym for the other words.

7.3.3. semantic association

After looking at the collocates of the words in query, we move on to the semantic associations. I classified semantic sets of the words in query based on their lexical category. Firstly I categorised adjective collocates of the words in query into two sets, namely Evaluation/Assessment (subcategorised into Positive, Neutral and Negative) and Logic. I then mapped these sets against the words in query, as in Table 7.8.

Words in query	Evaluation/Assessment			Logic
	Positive	Neutral	Negative	
影响 (*yíng xiǎng*) (influence)	深远 (*shēn yuǎn*) (profound) 积极 (*jī jí*) (active, positive) 大 (*dà*) (big) 深刻 (*shēn kè*) (incisive) 重要 (*zhòng yào*) (important), 巨大 (*jù dà*) (huge), 重大 (*zhòng dà*) (major)	广泛 (*guǎng fàn*) (extensive)	严重 (*yán zhòng*) (serious, critical)， 恶劣 (*è luè*) (adverse), 坏 (*huài*) (bad),	/
结果 (*jié guǒ*) (result)	/	/	/	/
成果 (*chéng guǒ*) (achievement)	丰硕 (*fēng shuò*) (plentiful)， 优秀 (*yōu xiù*) (excellent)， 新 (*xīn*) (new)， 好 (*hǎo*) (good)	/	/	/
效果 (*xiào guǒ*) (effect)	好 (*hǎo*) (good), 良好 (*liáng hǎo*) (fine)， 实实在在 (*shí shí zài zài*) (substantial)， 满意 (*mǎn yì*) (satisfactory), 显著 (*xiǎn zhù*) (notable, outstanding)， 理想 (*lí xiǎng*) (ideal)， 不错 (*bú cuò*) (not bad)， 佳 (*jiā*) (good, fine)	明显 (*míng xiǎn*) (obvious)，意想不到 (*yì xiǎng bú dào*) (unexpected)		/
后果 (hoù guǒ) (consequence)	/	/	严重 (*yán zhòng*) (serious, critical)， 不利 (*bú lì*) (unfavourable)， 可怕 (*kě pà*) (terrible)， 不良 (*bù liáng*) (not good)， 意想不到 (*yì xiǎng bú dào*) (unexpected)	直接 (*zhí jiē*) (direct)，
果实 (*guǒ shí*) (fruit)	丰硕 (*fēng shuò*) (plentiful)， 胜利 (*sheng lì*) (successful)，丰收 (*fēng shōu*) (harvestful)， 甜美 (*tián měi*) (sweet and nice)， 累累 (*lěi lěi*) (clusters of)， 成熟 (*chéng shú*) (ripe)， 红红 (*hóng hóng*) (red)， 沉甸甸 (*chén diān dian)*(heavy)， 甜蜜 (*tián mì*)(sweet)	/	恶魔 (*è mó*) (evil monster)	/

结局 (*jié jú*) (end)	圆满 (*yuán mǎn*) (satisfactory)， 完满 (*wán mǎn*) (satisfactory)， 完美 (*wán měi*) (perfect)	/	悲惨 (*bēi cǎn*) (miserable)， 出人意料 (*chū rén yì liào*) (unexpected)， 可悲 (*kě bēi*) (pitiable)， 悲凉 (*bēi liáng*) (sad)， 糟糕 (*zāo gāo*) (bad)， 坏 (*huài*) (bad), 凄惨 (*qī cǎn*) (miserable)， 无奈 (*wú nài*) (have to choice)	必然 (*bì rán*) (inevitable)
硕果 (*shuò guǒ*) (achievement)	累累 (*lěi lěi*) (clusters of)，丰收(*fēng shōu*) (harvest)，喜人 (*xǐ rén*) (gratifying)， 沉甸甸 (*chén diān dian*) (heavy)，丰厚 (*fēng hòu*) (rich), 丰盛 (*fēng shèng*) (rich, bumper)	/	/	/
恶果 (*è guǒ*) (evil fruit)	/	/	严重 (*yán zhòng*) (serious, critical)， 可怕 (*kě pà*) (terrible)	/
苦果 (*kǔ guǒ*) (bitter fruit)	/	/	/	/

Table 7.8 The semantic sets of adjective collocates associated with the words in query

There are two points worth mentioning here. First, 影响 (*yíng xiǎng*) (influence) and 效果 (*xiào guǒ*) (effect) share the semantic sets of Positive and Neutral Evaluation, but 效果 (*xiào guǒ*) (effect) does not have a semantic association with Negative Evaluation. In addition, although 影响 (*yíng xiǎng*) (influence) and 效果 (*xiào guǒ*) (effect) share the semantic sets of Positive Evaluation, the collocates of each one seem to emphasise different aspects of being positive. 影响 (*yíng xiǎng*) (influence) tends to have the positive association of DEPTH (for example, 深远 (*shēn yuǎn*) (profound) and 深刻 (*shēn kè*) (incisive)) and SIGNIFICANCE (for example, 重要 (*zhòng yào*) (important) and 重大 (*zhòng dà*) (major)), while 效果 (*xiào guǒ*) (effect) prefers positive associations of SATISFACTION (for example, 满意 (*mǎn yì*) (satisfactory) and 理想 (*lí xiǎng*) (ideal). 结局 (*jié jú*) (end) and 果实 (*guǒ shí*) (fruit) share both positive and negative associations, but 果实 (*guǒ shí*) (fruit) is primed more for positive than negative. Second, 成果 (*chéng guǒ*)(achievement) and 硕果 (*shuò guǒ*) (achievement) are only primed for positive association while 后果 (*hòu guǒ*) (consequence) and 恶果

(*è guǒ*) (evil fruit) are only primed for negative, which seems to suggest these four words are co-hyponyms and they and that we are looking at two pairs of synonyms.

Then I looked at the verb collocates of these words. A semantic set of CAUSE could be categorised comprising 酿成 (*niàng chéng*) (breed), 造成 (*zào chéng*) (cause), 带来 (*dài lái*) (bring about), 引起 (*yǐn qǐ*) (give rise to), and 产生 (*chǎn shēng*) (produce). Xiao and McEnery (2006) conducted a comparative analysis of English and Chinese *cause*-words. By looking at these words in the Lancaster Corpus of Mandarin Chinese corpus (LCMC) and the People's Daily (2000) Corpus for Chinese (PDC 2000), they discovered that these words differ in semantic association as shown in the following Table 7.9.

Synonyms	Negative	Positive	Neutral
酿成 (*niàng chéng*) (breed)	92 (98%)	2 (2%)	0
造成 (*zào chéng*) (cause)	190 (91%)	3 (2%)	15 (7%)
带来 (*dài lái*) (bring about)	64 (49%)	36 (27%)	31 (24%)
引起 (*yǐn qǐ*) (give rise to)	83 (43%)	28 (15%)	81 (42%)
产生 (*chǎn shēng*) (produce)	111 (31%)	88 (24%)	162 (45%)

Table 7.9 Semantic prosody of Chinese *Cause*-words (from Xiao and MeEnery, 2006)

As was discussed in Chapter 2, Louw (1993) defines semantic prosody as '[a] consistent aura of meaning with which a form is imbued by its collocates' and argues that the habitual collocates of a form are 'capable of colouring it, so it can no longer be seen in isolation from its semantic association'. Partington (1998) also emphasizes the spreading of connotation of single words through word boundaries, since semantic prosodies sometimes are interpretable in terms of connotations. If the node words and their collocates tend to co-occur frequently, they will acquire the same connotational features, and meaning and form merge.

7.3.4. colligation

There are no definite/indefinite articles in Chinese. Nor is there the concept of countability/uncountability of nouns. In consequence colligational analysis of the Chinese data has to make use of different grammatical categories from those often used for English. Before I present the result, a brief introduction to some grammatical terms in Chinese therefore seems necessary.

7.3.4.1. grammatical functions in Chinese

Chao (1968) defines subject as 'the first noun phrase' and predicate as 'the rest of the sentence'. Look at the following example:

这 个女孩 Zhèi ge nǚ hái	眼睛 yǎn jing	很大。 hěn dà.
This girl	eyes	very big
Subject	Predicate	

This girl has big eyes.

In this sentence, 这个女孩 (*zhèi ge nǚ hái*)(this girl) is the first noun phrase, hence the subject of the sentence. The rest of the sentence 眼睛很大 (*yǎn jing hěn dà*)(eyes very big) is the predicate.

A sample of 300 instances of each candidate word were retrieved from zhTenTen11 and the grammatical functions of each word in query in the clauses were analysed. Table 7. 10 shows the relative frequency (in percentage terms) of the grammatical distributions of the words in query in the clauses. The following findings deserve attention. Firstly, each word is primed with different strengths in terms of its grammatical position in the clause. For example, there is a positive colligation between 影响 (*yíng xiǎng*) (influence) and the grammatical function of Object as about 48% of instances of 影响 (*yíng xiǎng*) (influence) occur with this function. On the other hand, there is a negative colligation between 影响 (*yíng xiǎng*) (influence) and the function of Complement, with only 4.3% of instances serving this function. Secondly, among all the words in query, 结局 (*jié jú*) (end) is the most positively primed to occur with the function of Subject, and it is negatively primed to occur with the function of Adjunct. Thirdly, 苦果 (*kǔ guǒ*) (bitter fruit) is the most positively primed with the function of Object, but is the most negatively primed to occur with the function of Adjunct. Finally the category Others in Table 7.11 refers to those instances which appear as headings or individual lists of articles in the text, and all the words in query are negatively primed with this function, except for 硕果 (*shuò guǒ*) (achievement) which appears as the heading of a paragraph in text very often.

	Part of Subject	Part of Object	Part of Complement	Part of Adjunct	Others	Total
后果 (*hoù guǒ*)(consequence)	116 (38.7%)	116 (38.7%)	14 (4.7%)	40 (13.3%)	14 (4.7%)	300
影响 (*yíng xiǎng*)(influence)	78 (26%)	143 (47.7%)	13 (4.3%)	55 (18.3%)	11(3.7%)	300
恶果 (*è guǒ*)(evil fruit)	77 (25.7%)	136 (45.3%)	67 (22.3%)	9 (3%)	11 (3.7%)	300
成果 (*chéng guǒ*)(achievement)	61 (20.3%)	179 (59.7%)	18 (6%)	28 (9.3%)	14 (4.6%)	300
效果 (*xiào guǒ*) (effect)	87 (29%)	167 (55.7%)	12 (4%)	25 (8.3%)	9 (3%)	300
果实 (*guǒ shí*)(fruit)	116 (38.7%)	118 (39.3%)	20 (6.7%)	37 (12.3%)	9 (3%)	300

硕果 (*shuò guǒ*)(achievement)	54 (18%)	112 (37.3%)	33 (11%)	22 (7.3%)	79 (26.3%)	300
结局 (*jié jú*) (end)	120 (40%)	113 (37.7%)	35 (11.7%)	27 (9%)	5 (1.7%)	300
结果 (*jié guǒ*) (result)	115 (38.3%)	102 (34%)	36 (12%)	39 (13%)	8 (2.7%)	300
苦果 (*kǔ guǒ*) (bitter fruit)	34 (11.3%)	193 (64.3%)	43 (14.3%)	6 (2%)	24 (8%)	300

Table 7.10 A comparison of the grammatical distribution in the clause of the words in query

7.3.4.2. the identification of theme in Chinese

Now the analysis moves on to Theme and Rheme. Although the concept of Theme was put forward with examples from English, Halliday (1994/2000) has provided a general guide to the identification of Theme in other languages:

If in any given language the message is organized as a Theme-Rheme structure, and if the structure is expressed by the sequence in which the elements occur in the clause, then it seems natural that the position for the Theme should be at the beginning, rather than at the end or at some other specific point. (p. 38)

Chinese is a language in which utterances are organized in a Theme-Rheme structure, and the framework of SFG has been applied to the analysis of Chinese by several researchers (e.g., Fang, 1989; Fang & McDonald, 2001; Hu, 1997; Hu, 1994; Li, 2007). Dai (2009) notes that 'Theme in Chinese is generally identified as the left-most constituent of a clause, and is further classified into simple theme, clausal theme, and multiple theme' (p. 4). According to Halliday (1994/2000), a simple Theme is one that 'consists of just one structural element, and that element is represented by just one unit—one nominal group, adverbial group or prepositional phrase' (p. 39). Dai (2009) defines clausal Themes as appearing 'in clause complexes in which the clause in the initial position of the clause complex is identified as a clausal theme' and further explains 'both clausal themes and rhemes may further contain their respective themes and rhemes' (p. 4). Dai (2009) gives the following example to illustrate topical Themes, interpersonal Themes, and textual Themes in multiple Themes.

Example 7.18:

不过， 说 句 实话， 读 《家》、《春》、《秋》时 的 感情
Bú guò, shuō jù shí huà, dú jiā chūn qiū shí de gǎnqíng
but, tell CL truth, read *Family spring autumn* time MM empathy
不如 读 《平凡的 世界》 来得 强烈……
bùrú dú píngfán de shìjiè láide qiángliè
unequal-to read *Ordinary* MM *world* come strong
However, to tell the truth, my empathy for *The family, spring, and autumn* is not so strong as that for *The ordinary world*.'

不过 Bú guò	说句实话 shuō jù shí huà	读《家》、《春》、《秋》时的感情 dú jiā chūn qiū shí de gǎnqíng	不如读《平凡的世界》来得强烈 Bù rú dú píngfán de shìjiè láide qiángliè
however	to tell the truth	my empathy for *The family, spring, and autumn*	is not so strong as that for *The ordinary world*
Textual	Interpersonal	Topical	
Theme			Rheme

Table 7.11 Illustration of topical, interpersonal and textual themes in multiple themes

It has been mentioned (Chapter 5) that building on Halliday's work on Theme, Berry (1995, 1996) argues that Theme does not necessarily refer to only the first ideational element in a clause and argued that Theme should be extended up to and including the Subject. Following the approach adopted by Davis & Berry (1995; 1996) and also Hoey (1999), Li & Thompson (1989) has identified Theme as everything up to and including the subject or where there is no subject up to the predicate. They proposed a new approach to grammatical functions in Chinese and distinguished Topic from Subject and point out that the subject must always have a direct semantic relationship with the verb as the one that performs the action or exists in the state named by the verb, but the topic need not (p. 15). Look at the following example:

Example 7. 13

这棵树叶子很大。
Zhè kē shù yè zi hěn dà.
This CLtree leaf very big.
This tree, (its) leaves are very big. (Li and Thompson's example, 1989)

According to Li and Thompson (1989), the Subject in sentence 7.13 refers to the things that are very big, that is 叶子 (*yè zi*) (leaves) and the Topic is 这棵树 (*zhè kē shù*) (this tree), which does not have a direct semantic relationship with the verb.

The difference between the approaches to the analysis of Theme and Rheme in Chinese of Chao (1968) and Li and Thompson (1989) are shown in Table 7.12.

	这个女孩 *Zhèi ge nǚ hái*	眼睛 *yǎn jing*	很大。 *hěn dà.*
Chao's approach (1968)	Subject	Predicate	
	Topic	Comment	
	Theme	Rheme	
Li and Thompson's approach (1989)	Topic	Subject	Predicate
	Theme		Rheme

Table 7.12 Different approaches to analysis of Theme and Rheme in Chinese

I followed Li and Thompson's (1989) analytical approach in identifying Theme and Rheme and Table 7.13 shows the distributions of Theme/Rheme of the words in query in the clause. It can be seen that all the candidate words are more positively primed with the function of Rheme than with the function of Theme. Among all the candidate words, 苦果 (*kǔ guǒ*) (bitter fruit) has significantly positive primings with the function of Rheme, with the highest proportion of 86.7%. The word 结果 (*jié guǒ*) (result) has an almost even distribution in Theme and Rheme, with a proportion of 47% and 53% respectively.

	Theme	Rheme	Total
影响 (*yíng xiǎng*) (influence)	114 (38%)	186 (62%)	300
成果 (*chéng guǒ*) (achievement)	65 (21.7%)	235 (78.3%)	300
结果 (*jié guǒ*) (result)	141 (47%)	159 (53%)	300
效果 (*xiào guǒ*) (effect)	92 (30.7%)	208 (69.3%)	300
后果 (*hoù guǒ*) (consequence)	129 (43%)	171 (57%)	300
果实 (*guǒ shí*) (fruit)	121 (40.3%)	179 (59.7%)	300
结局 (*jié jú*) (end)	124 (41.3%)	176 (58.7%)	300
硕果 (*shuò guǒ*) (achievement)	55 (18.3%)	245 (81.7%)	300
恶果 (*è guǒ*) (evil fruit)	86 (28.7%)	114 (71.3%)	300
苦果 (*kǔ guǒ*) (bitter fruit)	40 (13.3%)	260 (86.7%)	300

Table 7.13 Distributions of Theme and Rheme of the words in query

Sentence structure is very loose in Chinese and clauses can be linked together only with commas and without any linking words. This means that the distinction between clause and sentence in Chinese is blurred. The following is an example of a very loosely constructed sentence:

> 《北京声明》表达了亚洲各国保护野生动植物资源、控制非法贸易、实现可持续发展的 心声和进行国际合作的愿望，是一份很好的文件，它将对亚洲地区的野生动植物保护事业产生深远的影响，是亚洲各国在控制野生动植物非法贸易方面开始携手并进的里程碑。
> The "Beijing Declaration" has expressed a desire for international cooperation among the Asian countries. These countries would work together to protect wildlife resources, control illegal trade, realize sustainable development. It is a very good document. It will have a significant influence on wildlife protection in Asia. It is a milestone that shows the Asian countries are making progress in controlling the illegal wildlife trade.

It is therefore impossible to distinguish sentence-initial or non-sentence-ininital clauses, my analysis only distinguishes Subject-Theme and non-Subject-Theme. Table 7.14 shows the distributions of Subject and Non-Subject-Theme of the words in query in the Theme.

	Subject Theme	Non-Subject Theme	Theme
影响 (*yíng xiǎng*) (influence)	78 (68.4%)	36 (31.6%)	114
成果 (*chéng guǒ*) (achievement)	61 (93.8%)	4 (6.2%)	65
结果 (*jié guǒ*) (result)	115 (81.6%)	26 (18.4%)	141
效果 (*xiào guǒ*) (effect)	87 (94.6%)	5 (5.4%)	92
后果 (*hoù guǒ*) (consequence)	116 (89.9%)	13 (10.1%)	129
果实 (*guǒ shí*) (fruit)	116 (95.9%)	5 (4.1%)	121
结局 (*jié jú*) (end)	120 (96.8%)	4 (3.2%)	124
硕果 (*shuò guǒ*) (achievement)	54 (98.1%)	1 (1.8%)	55
恶果 (*è guǒ*) (evil fruit)	77 (89.5%)	9 (10.5%)	86
苦果 (*kǔ guǒ*) (bitter fruit)	34 (85%)	6 (15%)	40

Table 7.14 Distributions of Subject and Adjunct of the words in query in the Theme

The following findings deserve attention. Firstly, all the candidate words are more positively primed with Subject Theme than with Non-Subject Theme. For example, 98.1% of 硕果 (*shuò guǒ*) (achievement) are primed for Subject Theme and only 1.8% occur with Non-Subject Theme. Secondly, although more positively primed for Subject Theme than for Non-Subject Theme, 影响 (*yíng xiǎng*) (influence) has the highest strength of colligation with non-Subject theme among all the words in query, with a proportion of over 30% of the total.

In conclusion, this section has looked at collocations, semantic associations and colligations of the Chinese candidate synonyms and the analysis shows that these words in query share primings and that the strength of similarity among these words can be measured. As with English, the similarities and differences among the candidate synonyms seem to suggest that some words are more synonymous than others, but it is still difficult to say some words are synonyms while others are not. Therefore although the notion of synonymy is also helpful for us to understand the relationships between Chinese words, it is difficult to distinguish between synonyms and non-synonyms. In some cases the boundary between synonyms and co-hyponyms is fuzzy.

7.4. Comparison between English and Chinese synonymy

The previous sections have tackled the related issue of the applicability of lexical priming categories to Chinese synonymy and the effectiveness of a corpus-driven approach to the identification and description of Chinese synonymy. By looking at the two sets of analyses, the following section will make a comparison between English and Chinese synonymy. Firstly, the analyses of the Chinese synonymous pair 同意 *(tóng yì)* and 赞同 *(zàn tóng)* and their English equivalent *AGREE* and

CONCUR, given in the previous chapter, will be compared. Secondly the analysis of the potentially synonymous *RESULT* group in Chapter 5 will be compared with the equivalent set of candidate synonyms grouped around 结果 (*jié guǒ*). Implications of the comparative studies will be also be addressed.

7.4.1. 同意 (*tóng yì*) vs. 赞同 (*zàn tóng*) and *AGREE* vs. *CONCUR*

To test the claim of lexical priming with Chinese synonyms, a pair of verbs 同意 (*tóng yì*) and 赞同 (*zàn tóng*) were chosen for the analysis. To better present the result English equivalent *AGREE* and *CONCUR* were also analysed and a comparative study between the two languages was conducted.

The comparative analysis seems to suggest that Chinese and English have a different range of synonyms; in other words, although a comparative study was successfully carried out between the Chinese synonymous pair 同意 (*tóng yì*) and 赞同 (*zàn tóng*) and the English synonymous pair *AGREE* vs. *CONCUR*, it does not follow that 同意 (*tóng yì*) is equivalent to *AGREE* or 赞同 (*zàn tóng*) to *CONCUR*, or visa versa. As the analysis shows that 同意 (*tóng yì*) and *AGREE* have different senses and although the two words share some collocates, there are many differences in their priming in terms of collocations, semantic associations and colligations in particular. Therefore I would argue that translating synonymous items from one language to another does not guarantee that the translated items are synonyms with the original items in the other language.

Partington (1998) and other authors question the concept of equivalence in translation. For example, Hatim and Mason (1990) points out:

There is also a problem concerning the use of the term 'equivalence' in connection with translations. It implies that complete equivalence is an achievable goal, as if there were such a thing as a formally or dynamically equivalent target-language version of a course language text. (p. 8)

Baker (1992) defines 'false friends' in translation as

> Words or expressions which have the same form in two or more languages but convey different meanings. They are often associated with historically or culturally related languages such as English, French and German. (p. 25)

Partington (1998) takes examples of two sets from English and Italian: (i) the ''look-alike'' items *correct* and *corretto*; (ii) the set of adverbs *absolutely*, *completely* and *entirely* and their counterparts *assolutamente*, *completamente* and *interamente*. His analyses show that 'even words of similar form considered to be excellent friends are not always reliable translation equivalents' (p. 63).

Chinese and English are from different language families and they are typographically different. We could not find out the same form in the two languages and the concept of 'false friends' seems not applicable. Baker argues that 'in fact false friends also abound among totally unrelated languages such as English, Japanese and Russian' (p. 25). The focus here is not on whether there are false friends

between English and Chinese. However the concept of translation equivalent seems to be relevant to distinction to whether we can distinguish synonyms in the two languages. A case in point is that the English words *efficient* and *effective* can both be translated into 有效的 (*yǒu xiào de*) in Chinese, which may be 有效果的 (*yǒu xiào guǒ de*, effective) and 有效率的 (*yǒu xiào lǜ de*, efficient). Although 有效果的 (*yǒu xiào guǒ de*, effective) and 有效率的 (*yǒu xiào lǜ de*, efficient) do not actually share the same meaning in Chinese either, the two English words *efficient* and *effective*, which are not considered as synonyms by any speaker of English, are often taught as synonyms to Chinese learners of English as they look similar in form to Chinese students and may be translated into similar expressions in Chinese. I would prefer to call this phenomenon as 'false friends' in the two unrelated languages.

7.4.2. 结果 (*jié guǒ*) and *RESULT* group

The preceding chapter conducted a corpus-driven study looking at a group of potentially synonymous English words and the analysis shows the complexity of the concept of synonymy. The words in the *RESULT* group can be synonyms, co-hyponyms, and even metaphors in different contexts. The current chapter analysed a group of Chinese potential synonyms all of which roughly mean *result*. The analysis shows that the words in query seem to form hyponymy, as 成果 (*chéng guǒ*) (achievement) and 硕果 (*shuò guǒ*) (achievement) are co-hyponyms of each other. Both meaning good 结果 (*jié guǒ*)(result), they are hyponyms of 结果 (*jié guǒ*)(result). Meanwhile, 恶果 (*è guǒ*) (evil fruit) and 苦果 (*kǔ guǒ*) (bitter fruit) are co-hyponyms of each other and also hyponyms of 结果 (*jié guǒ*)(result).

The comparative analysis conducted between the English and Chinese words in query may lead to the following implications. Firstly, different languages have a different range of synonyms and it should not be assumed that there is a complete translation equivalence of the members of candidate synonyms in any two languages. Secondly near-synonyms are normally not interchangeable in either language because of semantic association, a point which Xiao & McEnery (2006) also argue although they talk in terms of semantic prosodies and semantic preferences. Finally, although from different language families, the synonyms in the two languages share some similarities in terms of their collocational behaviours and semantic associations. In their comparison of the English synonyms *cause* group and the Chinese group of 造成 (zào chéng), Xiao and McEnery (2006) pointed out that 'both languages exhibit features of semantic prosody, but near synonyms are normally not collocationally interchangeable in either language as they show different semantic prosodies; consider *CAUSE* a change vs. *BRING* about a change, and 造成 (*zào chéng*) 结果 (*jié guǒ*) vs. 产生 (*chǎn sheng*) 结果 (*jié guǒ*)' (p. 120). This observation echoes the findings which have so far been reported for related language pairs, for example English vs. Portuguese (Sardinha, 2000), English vs. Italian (Tognini-Bonelli, 2001, p. 131–56), and English vs. German (Dodd, 2000).

Now we are in a position to answer the remaining research questions:

(3) Are the findings concerning synonyms derived from the analysis of Chinese data consistent with the findings concerning English synonyms derived from the same kind of analysis of English data? In other words, can we describe synonymy in the same way in both English and Chinese?

(4) If synonymy can be described in the same way in languages which have no family relationship, do the corpus-linguistic categories used by Lexical Priming enable us to identify similarities and differences between candidate synonyms in both English and Chinese?

(5) Given Cruse (2002)'s claims that synonymy is scalar, do the categories used in Lexical Priming help us to measure the strength of synonymy between pairs or among a group of words in the two unrelated languages?

Although English and Chinese do not have a family relationship, the corpus-linguistic used by Lexical Priming not only enable us identify the similarities and differences between candidate synonyms in both languages, but also help us to measure the strength of synonymy between pairs or among a group of words in the two unrelated languages.

7.5. Conclusion

Chapter 6 tested the applicability of lexical priming to Chinese and found that collocation, semantic association and colligation are observable in Chinese as they are in English. (See also Hoey & Shao 2015; Xiao & McEnery 2006). By conducting a case analysis of synonymous verbs 同意 (*tóng yì*) and 赞同 (*zàn tóng*) in Chinese, the claim concerning synonyms in lexical priming is supported; in other words, synonyms share collocations, semantic associations and colligations but differ in the strength of their distributions and proportions. The comparative analysis of English and Chinese synonymous pairs also shows that near synonyms and close translation equivalents in different languages may demonstrate, to some extent, different collocational behaviours and semantic associations.

This chapter looked at the concept of synonymy in Chinese using a data-driven approach. The results show a situation which is as complex as that in English. The concept of synonymy in Chinese is also not as straightforward as we expect. The candidate words can be in a superordinate-hyponym relation or co-hyponyms. For example, 结果 (*jié guǒ*)(result) can be considered as the superordinate of 苦果 (*kǔ guǒ*, bitter fruit), as 苦果 (*kǔ guǒ*, bitter fruit) is a bad 结果 (*jié guǒ*)(result). Also 恶果 (*è guǒ*)(evil fruit) and 苦果 (*kǔ guǒ*, bitter fruit) can be co-hypomyms, because being bad, they are both a type of 结果 (*jié guǒ*)(result).

Furthermore a number of implications can be drawn from the cross-linguistic study.

Firstly the reason that two distinctly unrelated languages share collocational behaviours and semantic

associations may be the ‘common basis of natural language semantics’ (Sweetser, 1990).

Secondly, different ways for people to conceptualise/classify their experiences may be the reason why different languages may have different ranges of near-synonyms. As well as emphasising the role of language as ‘an instrument of communication’, Leech (1981) pointed out ‘language is the means by which we interpret our environment, by which we classify or ‘conceptualize’ our experiences, by which we are able to ‘impose structure on reality’.

Although some of present-day thinking has tended to hypothesize a universal conceptual framework which is common to all human language, common observation shows that languages differ in the way they classify experience (Leech, 1981). A classic instance of this is the semantics of colour words. We have, like other creatures, the visual apparatus for discriminating colour differences. But in addition, unlike animals, we have the apparatus for categorizing these colours verbally. For example, English has a range of eleven main colour terms (‘black’, ‘white’, ‘red’, ‘green’, ‘yellow’, ’blue’, ‘brown’, ‘purple’, ‘pink’, ‘orange’, and ‘grey’) (Berlin and Kay, 1969), whereas Chinese language has six terms:

赤 chì -- red

橙 chéng -- orange

黄 huáng -- yellow

绿 lǜ -- green

青 qīng -- green-blue

蓝 lán -- blue

紫 zǐ -- purple

Biological or physical differences among ethnic groups are not the focus here. The different terms across languages show that different cultures classify or “conceptualize” our experiences in different ways.

In conclusion, the way people conceptualise their experiences plays a vital role in how people use their language. Meanwhile, languages play an important part in how people conceptualise their experiences.

Chapter 8 Concluding Remarks

This final chapter begins with the goal of the thesis and is followed by a summary of each chapter. The implications of the current research and recommendations for the future study are also included in the chapter.

8.1. Goals of the thesis

The subject of the thesis has been synonymy and there have been six overall goals. The first goal has been to explore the psychological reality of synonymy. The second goal has been to test whether corpus analysis of words traditionally considered to be synonyms justify retention of the concept of 'synonymy' and whether the findings arising out of the corpus approach and the findings of the psycholinguistic experiment match each other and support the notion of synonymy. The third goal has been to find out whether synonymy exists in the same way, or can be described in the same way, in languages which have no family relationship. The fourth goal has been to explore how a corpus linguistic approach is capable of showing similarities and differences between candidate synonyms in both English and Chinese. The fifth goal has been to investigate whether the categories in lexical priming assist us to measure the strength of synonymy between a pair of words or among a group of words in the two languages. The final goal has been to find out whether we can justify the continued use of the notion of synonymy based on the findings both from the psycholinguistic experiment and the corpus approach, and on what grounds.

8.2. Brief summary of each chapter

In addressing the six overall research goals, each chapter had its focus on specific issues. Chapter 1 introduced the motivation of the study and posed five overall research questions. Chapters 2 and 3 set up the scene for the study by reviewing the development of corpus linguistics and the literature of synonymy. Chapter 2 introduced key terms in corpus linguistics including collocation, colligation and semantic association, and also discussed some of the main contributions of corpus studies of synonymy. Chapter 3 began with definitions and classifications of synonymy, followed by a discussion of traditional approaches to identifying synonyms including substitution and componential analysis. The deficiencies of the two approaches were addressed. Although words may be offered as synonyms in dictionaries and thesauri, they may not always be substitutable for each other in different contexts; examples are *strong/*powerful tea* (Halliday, 1976) and *a big surprise/*a large surprise*. Componential analysis also was shown to be ineffective as it presupposed the feature of

'decomposing' words into their semantic components, which may be very difficult for words like *belief* and *faith*, which do not have component parts that can be enumerated. Then after a review of previous studies on synonymy in English and Chinese, the theory of lexical priming was put forward as the framework of the current study of synonymy.

Chapter 4 reported the findings of a psycholinguistic experiment on synonymy. A word association test was carried out to explore the psychological reality of synonymy. Thirty English words were chosen as the prompt words and forty-two subjects of four age groups from a local school near Liverpool and the University of Liverpool participated in the test. The results showed that people may not have a shared sense of synonymy. On the one hand, for different types of prompt words people may offer various words as candidate synonyms. The words provided by the participants may be considered on occasion to be co-hyponymous, metonymous, or meronymous or to be in a metaphorical relationship with the prompt words. On the other hand, there was found to be a relationship between candidate synonyms provided and the personal profile of the participants, including age, gender and subject field. Firstly it seems that older participants tended to provide a greater number and variety of putative synonyms than younger participants; secondly females tended to offer more synonyms than males; and finally the subject field of the participants may also have been a possible factor in deciding which words to offer as synonyms. This may be caused by the various contexts and for which people are primed. The experiment result also showed there is a scale of similarity among the candidate synonymous words provided by the participants.

Chapters 5 analysed a group of English candidate synonyms, comprising *result, outcome, aftermath, upshot, sequel, effect, end-product, by-product, fruit, impact* and *consequence*. The results of this analysis seemed to suggest that the similarities and differences among the candidate synonyms could be demonstrated in terms of their primings with collocations, colligations and semantic associations and that this corpus linguistic approach may enable us to measure the strength of a group of candidate synonyms. In addition, strength of similarities and differences among candidate synonyms in terms of their collocations, semantic associations and colligations may suggest the closeness of meanings. However the analysis did not answer the question whether two candidate words are synonyms. As it was found that there is a scale of similarities in terms of meaning/sense, we can only say that two words are highly synonymous or synonymous to a certain degree, but we cannot decide at what point on the scale two words become synonyms or not. It was therefore concluded that traditional semantic relations such as synonymy, antonymy, hyponymy, metonymy and meronomy could only help us understand how words may be related to each other, but that it was not always possible to allocate a pair of words to one relation rather than another. This finding seems to be consistent with what we found in the psycholinguistic experiment. Combining the findings both from the psycholinguistic experiment and corpus analysis, we believe we have shown that the notion of synonymy is more

complex than we may think and that the ways people are primed may suggest possible explanations for the complexity of this linguistic phenomenon.

Chapters 6 and 7 looked at synonymy from a cross-linguistic perspective. To identify whether the corpus-linguistic categories used by Lexical Priming enable us to identify similarities and differences between Chinese candidate synonyms, it was necessary to test the applicability of the theory of Lexical Priming to Chinese. In Chapter 6 a pair of synonymous verbs 同意 (*tóng yì*) and 赞同 (*zàn tóng*) were chosen for a case study. The results showed that the categories utilised in lexical priming such as collocation, semantic association and colligation could be observed in Chinese. Then a comparative analysis of the verbs and their English equivalents *AGREE* and *CONCUR* was conducted. The comparative analysis seemed to show that the Chinese synonymous verbs 同意 (*tóng yì*) and 赞同 (*zàn tóng*) are more divergent than the English pair. One possible explanation may be that the English pair is more closely synonymous than the Chinese pair. The deficiencies of the Chinese corpus however could not be ignored as a possible explanation.

As the analysis of the Chinese synonymous pair 同意 (*tóng yì*) and 赞同 (*zàn tóng*) showed that they do not share as many primings as we might expect of candidate synonyms, one might argue that 同意 (*tóng yì*) and 赞同 (*zàn tóng*) are not synonymous. This seemed to raise the question whether synonymy exists in Chinese in a form identifiable by the categories used to identify in English. Therefore it was necessary to look at more data in Chinese within the framework of lexical priming. To achieve this purpose, in Chapter 7 a group of potentially synonymous words was analysed, namely 影响 (*yíng xiǎng*)(influence), 成果 (*chéng guǒ*)(achievement), 结果 (*jié guǒ*)(result), 效果 (*xiào guǒ*) (effect), 后果 (*hoù guǒ*)(consequence), 果实 (*guǒ shí*)(fruit), 结局 (*jié jú*)(end), 硕果 (*shuò guǒ*) (achievement), 恶果 (*è guǒ*)(evil fruit) and 苦果 (*kǔ guǒ*)(bitter fruit). The results showed that these Chinese candidate synonyms do share primings in terms of collocation, semantic association and colligation but they differ in the strength of their distributions. In addition, the analysis of Chinese data seemed also to suggest that shared primings could only help us demonstrate the closeness of meanings between candidate words but not decide whether they are synonyms or not. As with English, the boundary between synonymy and co-hyponymy may be blurred. Then the analyses of the eleven English candidate synonyms (Chapter 5) and ten Chinese candidate synonyms (Chapter 7) were compared and this showed that both English and Chinese synonymy could be described in the same way, that is, the features of synonymy could be characterised in terms of the same categories utilised in lexical priming, namely collocation, semantic association and colligation. The differences between English and Chinese synonyms were differences in their own primings with collocations, semantic associations and colligations. In other words, the fact that words are synonyms in one language does not guarantee that their nearest equivalents will be synonyms in the other language.

8.3. Implications of the study

Traditional linguistics describes language as systematic and well-organised. However language in real life can be messy and chaotic. To have a better understanding of what language really is, it is time to look at language from a different approach. With the development of computer technology, a quantitative approach to language is adopted in many branches of linguistics. By using a large number of authentic language data, corpus linguistics has attested and also disputed traditional language descriptions. Language is under a 'telescope' and starts to reveal its true features. Recent linguistic studies have given us some cases where looking at real-life data has produced modifications to the way we think about language (for example, Sinclair, 1991; Hoey, 2007).

Grammatical categories prove to be not clear-cut; meaning and form are not as separated from each other as we thought. These findings seem to call for the re-thinking of other linguistic concepts.

The term synonymy has been accepted and used for a long time; it has served its value in helping us understand semantic relations of lexemes in a systematic way. The current study has revisited the notion of synonymy and conducted a corpus-driven analysis of potentially synonymous items both in English and Chinese. Combining the findings of both the corpus approach and the psycholinguistic experiment, this research has explored ways of re-characterising the features of synonymy. It is difficult to distinguish synonymy and non-synonymy, but lexical priming provides a possible way of accounting for synonymy and for the difficulty in distinguishing synonymy from non-synonymy.

The research has explored the characterisations of synonymy from both lexical and psycholinguistic perspectives and the findings have both theoretical and methodological implications and also applications in translation and pedagogy.

8.3.1. theoretical implications

This research has implications for the theory of lexical priming. Firstly, within the framework of lexical priming, this study has shown that the corpus linguistic categories utilised by lexical priming can help identify similarities and differences between candidate synonyms in both English and Chinese. It not only supports the claim that lexical priming is not culture or language specific, but also demonstrates that synonymy can be described in the same way in two languages which do not have any family relations. Secondly, Hoey's (2005) theory of lexical priming can be criticised on the basis of a lack of psycholinguistic research (Williams, 2006). This study has attempted to address this criticism by presenting the finding of a psycholinguistic experiment. Hoey drew an analogy between what he thought of as a mental concordance and the traditional corpus concordance and pointed out

> the computer corpus cannot tell us what primings are present for any language user, but it can indicate the kind of data a language user might encounter in the course of being primed. It can suggest the ways in which priming might occur and the kind of feature for which words or word sequence might be primed. (2005, p. 14)

The result of the psycholinguistic experiment conducted in the current research is consistent with that of the corpus analysis, which supports Hoey's claim.

8.3.2. methodological issues

Although the corpus approach adopted has proved effective in exploring the linguistic characteristics of candidate synonyms, some methodological issues related to the current study need to be addressed. The first issue concerns statistics. Based on quantitative data, corpus approach works on the 'distributional hypothesis', which Gries (2015) defines as 'the working assumption that linguistic elements that are similar in terms of their distributional patterning in corpora also exhibit some semantic or functional similarity' (p. 94). The notions of collocation and co-occurrences in corpus linguistics are statistically based; frequencies of lexical items are the basis for quantitative analysis. Since raw frequencies can be distorted by highly frequent words, a popular strategy is to use association measures (Gries, 2015). However a couple of statistical issues about association measures have been raised, for example, 'the impact that dispersion, type frequencies/ entropies and directionality (should) have on the computation of association measures …' (see Gries, 2015 for details), and therefore the quantitative approach to language need to be further refined and developed.

Secondly the comparability of corpora is very important for cross-linguistic studies. As Aijmer and Altenberg (1996) observe, parallel and comparable corpora can 'offer specific uses and possibilities' (p. 12) for contrastive studies. McEnery at al. (2006) make a distinction between parallel and comparable corpora and point out that 'parallel corpora are undoubtedly a useful starting point for contrastive research, which may lead to further research in contrastive studies based upon comparable corpora' (p. 95). Many large-scale English corpora are available, but there are relatively fewer Chinese corpora. Although substantial progress has been made on building Chinese corpora since 1990s in Mainland China, there are not so many practical parallel corpora with Chinese involved available. The current cross-linguistic research into synonymy has been based on the BNC corpus and the Chinese zhTenTen11 corpus, which are both general corpora of considerable size. However as Chinese zhTenTen11 is a web-crawling corpus, it is difficult to trace the source of data and therefore almost impossible to know the genre or text type of each instance. Since the current study focuses on the general characteristics of near-synonymy rather than on genre-specific descriptions, the findings are unlikely to be affected. But features of synonymy concerning genre or text types might have been further provided if there had existed a Chinese corpus which had a similar construction to that of the

BNC corpus or if there had been more practical comparable corpora with English and Chinese involved available.

Thirdly work on annotation of Chinese data is far behind that of English. As mentioned in Chapter 1, grammatical categorisation in Chinese is more complicated than that in English. In addition automatic segmentation of Chinese is very demanding. More work is needed in these fields. Finally this study utilises the Sketch Engine, one of the few language analysis tools which is capable of analysing both English and Chinese data. The sketch engine has proved its value in analysing Chinese data and contributed to insightful findings. However, it does not seem to be effective all the time. For example, in the analysis of Chinese candidate synonyms, the sketch engine could only elicit two categories of collocates – modifiers and modifieds. The result may be due to the issues of grammatical categorisation and automatic segmentation in Chinese. It is necessary to develop language analysis software which can be more effective for Chinese data.

8.3.3. applications in pedagogy and translation

The implications of the study in translation and language teaching, particularly second/foreign language teaching, cannot be neglected. First of all, word choice from a group of apparently synonymous items for the appropriate context can be very challenging for second/foreign language learners. As an English teacher in China, I have noticed that recognising and distinguishing synonyms is very important and difficult for English language learners. Once one of my students in college was talking about his experience of entering a competition and he used the following words:

* I defeated all the difficulties and won the first prize in the competition.

When I pointed out that the word defeat was not properly used in the sentence and suggested that he use *overcome*, he was very confused and argued that he wanted to express 战胜困难 (*zhàn shèng kùn nán*, overcome difficulties). He remembered that 战胜 (*zhàn shèng*) meant *defeat* in English and had checked its definition in a dictionary. His dictionary was in fact not very helpful as it only provided a translation along with one or two examples. This was not an isolated case but happened very often in many years of English teaching. Similar situations also occurred in my later experiences as a Mandarin Chinese teacher in UK, which made me realise that such incomplete equivalences might be one of the reasons that students make mistakes in word choice.

In fact the problem exists not only for second/foreign language learners, but also for first language learners. According to Higa (1963), pairs of synonyms take longer to learn than pairs of unrelated words and learners are more likely to confuse words that are similar in meaning than words that do not have close semantic links. Studies by Tinkham (1993) and Waring (1997) also indicate that learning sets of semantically related words is more difficult than learning words that are not linked by

meaning. Dictionaries and thesauri may not be helpful either as they often offer a list of synonyms without giving information on their different collocational and colligational behaviours, which may lead to frustration and further confusion when consulted. Language teachers or practitioners have to rely on their intuition and introspection to help students with difficulties in word choice among synonyms, and these have been shown by corpus linguistics to be unreliable (see for example Stubbs, 1995). Therefore it seems important to find ways of helping students to choose the appropriate word for the right context in their speaking and writing. Although the traditional prescriptive approach to language does not help the learner distinguish between synonyms in language teaching, corpus methods have been shown in this thesis to be effective ways to differentiate synonyms both in English and Chinese. Therefore this kind of research can be applied in language teaching, particularly with respect to the distinction between synonyms in the two languages. Corpus-based dictionaries of synonyms should be compiled as they may present more accurate information about the similarities and differences of synonyms in terms of their linguistic behaviours in collocations, semantic associations and colligations, all of which are vital factors in deciding the word choice for both language learners and translation practitioners.

In addition the experience of learning L1 is very different from L2. As Hoey (2005) mentions

> When the vocabulary of the first language is primed, it is being primed for the first time. When the second language is learnt, however, the primings are necessarily superimposed on the primings of the first language. (p. 183)

Typically, in second/foreign language learning, L2 learners use their first language (L1) knowledge of that item and information from the context in which it was encountered to help learn that word. Due to the different primings between a word in one language and its translation in another, students may be confused about the appropriate meaning/senses and usage of the lexemes in the second/foreign language. Apparent synonymy may complicate matters further. Take *strong* and *powerful* and their Chinese translated equivalents 强壮的 (*qiáng zhuàng de*) and 强劲的 (*qiáng jìng de*) for example. The two nearly synonymous pairs are divergent (behave) differently in English and Chinese. We say *strong tea* and *powerful engine* in English. In Chinese it is acceptable to say 强劲的马达 (*qiáng jìng de mǎ dá*, powerful engine) but never *强壮的茶 (*qiáng zhuàng de chá*). Due to differences among the synonymous items in two languages, second/foreign language learners tend to make more mistakes in using synonyms. The priming transfer from Ll to L2 may be unavoidable ‘except where the learner learns through immersion and is never tempted by word-for-word translation’ (Hoey, 2005, p. 183); therefore awareness raising of different primings between L1 and its translated equivalent in L2 and classroom instruction on distinguishing these differences of synonyms between L1 and L2 should be emphasised in second/foreign language teaching.

8.4. Limitation and recommendations for the future study

Synonymy is a context-bound phenomenon. Firth (1957) emphasises the importance of context when he states, ‘no study of meaning apart from a complete context can be taken seriously’ (p. 7). W. E. Collinson (1939) suggests that ‘one must never study synonyms as isolated items, but must always study their functions when they are embedded in suitable contexts and figure in clearly apprehended situations’ (p. 58).

In addition, cohesion deals with the meaning in text and it is achieved by the ‘selection of vocabulary’ and it concerns the way in which ‘lexical items relate to each other and to other cohesive devices so that textual continuity created’ (Halliday & Hasan, 1976). It can be realised through ‘the repetition of a lexical item, at one end of the scale; the use of a general word to refer back to a lexical item, at the other end of the scale; and a number of things in between – the use of a synonym, near-synonym, or superordinate’(Halliday & Hasan, 1976).

This thesis however did not address the contextual aspect of synonymy and the relationships between cohesion and synonymy, which would be recommended for future studies.

Appendixx: The Complete Responses for all the Prompt Words

Number	Prompt word	Elicited words
1	amazing	brilliance (2), extraordinary (3), good (10), fantastic (24), awesome (8), incredible(7), brilliant (23), super (5), great (16), fabulous (12), powerful (1), wonderful (13), astonishing (4), stunning (2), unbelievable (4), superb (4), terrific (2), stupendous (2), excellent (7), smashing (1), exceptional (1), startling (1), shocking (1), perfect (2), magical (1), cool (2), magnificent (1), tremendous (2), happy (1), delightful (1), wow (3), exciting (1), formidable (1), special (1), nice (1), beautiful (1), lovely (1), unique (1), unreal (1), spectacular (1)
2	brave	willing (2), fearless (14), strong (13), confident (4), bold (5), force (1), courage (2), courageous (24), noble (3), fierce (3), heroic (6), determined (1), resilient (1), daring (1), valiant (1), valorous (1), adventurous (2), powerful (1), stupendous (1), unafraid (1), foolish (1), gallant (1), knightly (1), forceful (1), hard (1), focused (1), secure (1), challenged (1), selfless (1), hero (1), soldier (1), loyal (1)
3	famous	noted (1), notorious (2), recognised (1), celebrated (1), starring (1), rich (1), starry (1), star (6), film star (1), renowned (2), stardom (1), important (1), renowned (2), infamous (3), recognisable (1), familiar (1), legend (2), icon (2), aware (1)
4	happy	fun (3), over-joyed (4), joy (2), cheerful (13), joyful (11), smily (7), bubbly (1), excited (6), carefree (1), upbeat (3), gleeful (1), glad (4), smiling (6), laughing (1), ecstatic (6), jovial (2), good-natured (1), satisfied (2), replete (1), humorous (1), good-willed (1), joyous (8), thrilled (2), jolly (3), elated (2), marvelous (1), soulful (1), laughing (1), singing (1), smile (2), gleaming (1), delighted (4), blissful (1), relaxed up (1), chipper (1), positive (5), pleased (6), merry (2), content (5), contented (1), overwhelmed (1), cheery (1), alive (1)
5	neat	tidy (40), clean (19), presented (1), smart (4), clear (2), organized (14), well presented (3), orderly (2), smart (1), careful (3), well-ordered (1), orderly (1), cool (2), brilliant(1), level (2), sorted (1), particular (1), collected (1), ordered (3), well-organized (2), sharp (1), straight (1), arranged (1), compact (1), good (1), brimmed (1), strategic (1), logical (1), uniformed (1), regimented (1), routine (1), in order (1), formal (1)
6	strange	wired (34), funny looking (1), different (15), crazy (1), unknown (5), suspicious (1), unusual (23), mysterious (1), unordinary (1), not normal (2), peculiar (8), eerie (1), odd (15), sinister (1), scary (1), discomforting (1), bizarre (3), extraordinary (1), foreign (1), alien (2), new (2), queer (3), abnormal (3), spooky (1), awful (1), unique (3), silly (1), out of ordinary (1), unnatural (2), eccentric (3), unknown (2), unfamiliar (2), quirky (2), unexpected (3), alternative (1), disturbing (1), whacky (1), unconventional (1), awkward (1)

7	true	believe (1), belief (1), fact (7), correct (27), honest (15), right (13), truthful (3), loyal (4), faithful (2), fair (1), level (1), coordinated (1), measured (1), honesty (1), wholesome (1), kind (1), accurate (9), proved (1), known (1), factual (3), real (4), affirmative (1), valid (2), unquestionable (1), unarguable (1), trusted (1), expected (1), absolute (1), positive (1), certain (1), genuine (1), exact (1), verify (1)
8	calm	quiet (14), relaxed (17), chilled-out (2), non-nervous (1), easy going (1), peaceful (19), tranquil (9), chilled (3), unstressed (2), at ease (3), sleepy (1), untroubled (1), still (6), measured (1), restful (4), remote (1), soothing (2), balmy (1), soft (1), collected (2), motionless (1), sooth (1), slow (1), relaxing (1), serene (5), staid (1), sedate (1), passive (1), patient (1), content (1), placid (2), smooth (1), zen (1), rest (1), static (1), mellow (1), cool (1), silent (1), warm (1), subdued (1), laid back (1), mindful (1), steady (1)
9	fair	even (10), equal (13), same (1), balanced (8), sharing (1), king (1), helpful (1), both sided (1), agree (1), unbiased (3), just (10), open-minded (1), honest (5), true (2), pale (2), blond (4), proper (1), carnival (1), light (6), right (6), good (2), open (1), 50:50 (1), beautiful (1), correct (1), accurate (2), judge (1), mild (2), pretty (1), pleasing (1), fete (1), fairground (1), consistent (1), moral (2), accepted (1), reasonable (1), justified (1)
10	quiet	shy (7), silent (28), peace (1), peaceful (12), calm (11), echoless (1), calming (2), timid (3), whisper (8), sleepless (1), relaxed (1), reserved (1), tranquil (3), sleepy (2), untroubled (1), soundless (1), still (3), thoughtful (1), perceive (1), noiseless (5), subdued (1), soft (1), gentle (1), content (1), silence (1), shush (1), hushed (3), low (3), retiring (1), slow (1), inaudible (1)
11	difference	else (1), change (8), opposite (6), different (1), unusual (1), unique (4), extraordinary (1), abstract (1), diverse (1), strangeness (1), unmatched (1), oddness (1), variety (1), contrary (2), distinction (1), compare (1), alternative (2), odd (2), subtract (4), not the same (2), changed (1), unlike (1), contrast (4), dissimilar (2), variation (1), opposing (1), companion (1), variety (1), diversity (1), equivalence (1), not alike (1), not similar (1)
12	fruit	orange (2), apple (3), exotic (1), pineapple (1), vegetable (2), food (5), veg (2), healthy (3), vitamins (1), vegetarian (1), produce (4), result (1), reward (1), seed (3), bud (1), ripe (1), vegetation (1), pip (1), snack (1), natural (1), harvest (1), orchard (1), fresh (1), offspring (1)
13	fear	scared (26), worried (5), horror (7), panicked (1), nervous (4), frightened (7), terrified (6), unconfident (1), frightful (2), phobia (4), fright (5), petrified (2), anxious (4), dread (3), discomfort (1), unsettlement (1), anxiety (7), scare (6), terror (7), shock (1), worry (3), apprehensive(1), apprehension (2), alarmed (1), fearful (1), stressed (1), threatful (1), petrified (1), terrify (1), afraid (4), panic (2), frighten (1), nervousness (1), timid (1), trepidation (1), wary (1),

		unexpected (1), challenged (1), unexplained (1), harm (1), unease (1), extreme (1), chilled (1), danger (1)
14	idea	thinking (1), impossible (1), brainstorm (3), thought (31), opinion (2), source (1), plan (9), change (1), notion (2), theory (2), consideration (1), outline (1), clue (2), reason (1), explanation (1), suggestion (3), concept (4), brainwave (1), brainstorm (1), spark (1), enlighten (1), motive (1), conclusion (1), inspiration (3), conjecture (1), invention (3), inkling (1), innovation (1), point (1), solution (1), topic (1), subject (1), knowledge (1), rationale (1), feeling (1), instinct (1), conception (1), thesis (1), hypothesis (1), inspire (1)
15	trouble	naughty (9), danger (4), hazard (1), toxic (1), panic (1), struck (1), non approachable (1), hard (1), mistake (1), mischievous (1), despicable (1), mean (1), bad situation (1), bad (7), need help or assistance (1), unrest (2), disturbance (1), disruption (1), strife (2), woe(1), issue (3), concern (1), problem (7), ailment (1), bother (1), worry (1), thought (1), disquiet (1), mischief (3), chaos (1), fear (1), torment (1), turmoil (1), unhelpful (1), mayhem (1), naughtiness (1), situation (1), adventurous (1), difficulty (3), inconvenience (1), war (1), discord (1), unexpected (1), disrespect (1), fight (1), conflict (1)
16	consequence	risk (1), action (8), result of (25), punishment (4), payback (1), discipline (1), outcomes (9), change (1), after affect (1), getting what you deserve (1), end-product (1), response (1), issue (1), follow-on (1), repercussion (2), reaction (1), ramifications (3), sanction (2), effects (4), outcomes (1), as a result (1), because (of) (2), end-point (1), scenario (1), by-product (2), conclusion (2), resulting (1), happening (1), cause (1)
17	place	somewhere (4), else (1), different (1), culture (1), location (25), area (21), GPS (1), building (2), land (1), country (1), continent (1), anywhere (1), site (2), region (3), setting (1), put (6), resort (1), district (1), town (2), city (2), village (2), country (1), locate (2), square (1), center (1), position (4), zone (4), home (3), structure (1), space (2), destination (3), settlement (1), spot (2), situation (2), put down (1), drop (2), vicinity (1), niche (1)
18	story	book (7), writing (1), reading (1), tale (28), fairytale (1), novel (4), creation (2), legend (4), myth (3), background (1), meaning (1), fiction (7), poem (2), fable (12), adventure (1), parable (2), ode (2), yarn (2), lie (1), anecdote (4), idea (2), plot (3), history (1), gossip (1), narrative (9), prose (1), text (1), falsehood (1), article (1), recap (1), account (1), frication (1), retelling (1), floor (1), folklore (1)
19	by-product	item (2), material (1), thing (1), derivative (1), result (10), consequence (7), issue (1), follow-on (1), spin-off (3), waste (6), residue (2), off-shoot (1), adaptation (1), off-spin (1), outcome (4),

		secondary effect (1), remainder (1), leftover (1), bonus (2), extra (2), resultant (1), because (1), ramification (1), situation (1), side-effect (1), addition (1), residual (1), end result (1),
20	answer	question (2), reply (11), right (1), wrong (1), opinion (1), thought (1), decide (1), result (16), guess (1), respond (4), fact (3), opinion (1), statement (1), correct (1), response (10), reaction (2), retort (2), solution (12), ideas (1), outcome (2), proof (1), sum (2), feedback (1), conclusion (3), consequence (1), desire (1), speak (1)
21	begin	start (42), go (8), commence (16), fresh (1), renew (1), create (1), make (1), first (4), firstly (3), outset (1), get going (2), introduce (1), first movement (1), introduction (1), open (1), off (1), initial (2), kickoff (2), proceed (1), embark (1), opening (1), birth (1), open (1), initiate (3), end (1)
22	cry	sad (10), tears (4), upset (9), tear (3), melancholy (1), weep (14), sob (18), destroyed (1), tearful (9), tearing (1), wail (8), howl (4), yell (2), scream (5), shout (8), snivel (1), loment (1), call (2), shed tears (3), bawl (6), sorrow (2), moan (6), suffer (1), whinge (6), whine (1), blubber (4), whimper (3), teary (2), emotional (2), shriek (3), blub (1), well up (1), breakdown (1), yelp (1), holler (1), bellow (1), emotion (1), emote (1), release (1), shed (1), giggle (1)
23	decide	decision (7), choose (20), chosen (1), select (4), opinion (1), pick (10), answer (2), result (3), choice (5), making your mind up (1), resolve (2), determine (1), do (2), conclude (4), deduce (2), step (1), calculate (1), settle (1), finalise (1), next (1), conclusion (1), agree (1), settle on (1), weigh(-up) (2), accept (1), consider (2), evaluate (2), judge (2), think (1), strategy (1), predict (1), outcome (1), work out (1)
24	agree	sorting (1), deciding (1), decided (1), correct (2), right (1), my opinion (1), same (1), fair (2), accept (6), along (1), decision (1), concur (14), harmonise (1), consent (1), affirm (2), comply (5), nod (2), approve (2), yes (2), also (1), add (1), same (1), complicit (1), acknowledge (1), confirm (6), unite (1), match (1), sympathise (1), coincide (1), acquiesce (1), adhere (1)
25	describe	talk about (3), explain (9), tell (7), story (1), explain (4), inform (2), show (3), exaggerate (1), illustrate (6), depict (5), outline (1), relate (1), narrate (2), draw (1), list (1), break down (2), mean (1), display (1), analyse (1), paint (2), portray (2), recall (2), illuminate (1), articulate (1), say (2), resist (1), remember (1), elaborate (1), elucidate (1), picture (1), portrait (1), detail (1), communicate (1),
26	explain	telling (1), answer (4), tell (12), describe (17), share (1), understand (1), understanding (1), what is it (1), illustrate (2), relate (1), narrate (2), list (1), justify (3), mean (1), guidance (1), guide (1), show (3), reason (3), evaluate (1), talk (1), details (2), inform (4), outline (1), clear up (2), clarify (1), elaborate (3), articulate (1), expand on (1), mention (1), offer (1), recall (1), educate (1), enlighten (1), communicate (1), break down (1)

27	help	helped (1), helping (1), polite (1), kind (1), teamwork (2), partnership (1), friendly (1), nice (1), SOS (3), struggle (1), struck (1), rescue (2), solve a problem (1), aid someone (2), save (3), assistance (4), urgency (1), assist (23), aid (18), succor(1), co-operate (1), cooperation /(1), respond (1), support (8), comfort(1), sustain (1), care (2), guide (2), guidance (1), provide (1), relieve (1),
28	look	seeing around (1), watch (4), see (23), observe (9), sight (3), stare (14), focus on (4), peer (3), scan (1), search (3), hunt (1), saw (1), stared (1), gaze (5), fixate (1), glance (9), seek (2), squint(1), pry (1), visualise (2), check (2), view (8), perspective (2), reflect (1), regard (3), perceive (1), acknowledge (1), eye sight (1), notice (2), glare (3), gape (1), peep (2), peek (2), glimpse (1), browse (2), appearance (1), style (2), fashion (2), pursue (1), consider (1), vision (1), optical (1), examine (1), experience (1), oversee (1)
29	plan	think (2), blueprint (2), ideas (1), brainstorm (1), overview (1), decision (2), thought (1), idea (3), description what is to happen (1), detail (1), prepare (5), preparation (1), thoughts (1), outline (2), arrange (6), organise (9), suggestion (1), design (1), organize (3), drawing (1), list (2), plot (3), to do (2), decide (4), foresight (1), map-out (2), revise (2), layout (3), sort out (1), scheme (2), map (4), strategy (3), predict (1), structure (1), construct (1), formulate (1)
30	accept	loss (1), win (1), let (1), understand (1), live with (2), agree (17), fair (2), along (1), positive (1), okay with something (1), comes to terms with (1), understanding (1), embrace it (1), collect (1), acknowledge (3), take on board (1), allow (5), take (10), concur (2), condone (1), ignore (1), settle (1), have (1), keep (1), want (1), hold on (1), willing (1), correct (1), confirm (2), receive (4), welcome (1), support (1), admit (1), acquire (1), comply (1), concede (1)

Bibliography

Aristotle. (2014) 'Categories', in Jonathan, B. *The Complete Works of Aristotle, 2 vols*. Transl. Ackrill, J. L. Princeton: Princeton University Press, pp. 2510.

Austin, J. (1962) *How to do things with words*. Oxford: Clarendon Press.

Baker. M. (1992) *In Other Words: A Course book on Translation*. London and New York: Routledge.

Bawcom, L. (2010) *What's in a name?: The Functions of similonyms and their lexical priming for frequency*. Unpublished thesis. University of Liverpool.

Belfarhi, K. (2013) 'The componential analysis of literary meaning', *Colombian Applied Linguistics Journal*, 15(2). Bogotá, Colombia, pp. 288 - 301.

Berlin, B. and Kay, P. (1969) *Basic Color Terms*. University of California Press, Berkeley.

Biber, D., Conrad, S., and Reppen, R. (1998) *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Bolinger, D. (1975) *Aspects of language.* 2nd edn. New York: Harcourt Brace Jovanovich.

Borchert, D. M. (2006). *Encyclopedia of Philosophy*. 2nd edn. Detroit: Thomson Gale

Brownlow, S., Rosamon, J. A., & Parker, J. A. (2003) 'Gender-linked linguistic behavior in television interviews', *Sex Roles*, 49, pp. 121–132.

Bursill-Hall, G. L. (1960) 'The linguistic theories of J. R. Firth', in *Thought from the Learned Societies of Canada, Toronto*, pp. 237–50.

Chang, Y. C., Chang, J. S., Chen, H. J., & Liou, H. C. (2008) 'An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology', *Computer Assisted Language Learning*, 21(3), pp. 283-299.

Chao, Y.-R. (1968) *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California press.

Chief, L.-C., Huang, C.-R., Chen, K.-J., Tsai, M.-C. and Chang, L.-l. (2000) 'What Can Near-Synonyms Tell Us?', *Computational Linguistics and Chinese Language Processing* 5(1), pp. 47-60.

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.

Clark, H. (1970) 'Word Associations and Linguistic Theory', in Lyons, J. (ed.) *New Horizons in Linguistics.* Penguin, pp. 271-186.

Colley, A., Todd, Z., Bland, M., Holmes, M., Khanom, M., & Pike, H. (2004) 'Style and content in emails and letters to male and female friends', *Journal of Language and Social Psychology*, 23, pp. 369–378.

Collinson, W. E. (1939) 'Comparative synonyms: some principles and illustrations', *Transactions of the Philological Society*, 38 (1), pp. 54-77.

Cruse, D. A. (1986). *Lexical semantics.* Cambridge: Cambridge University Press.

Cruse, D. A. (2002) 'Paradigmatic relation of inclusion and identity III: Synonymy', in Cruse, A., Hundsnurscher, F., Job, M. and Lutzeier, P. R. (eds) *Lexicology--A international handbook on the*

nature and structure of words and vocabularies. Berlin: Walter de Gruyter, pp. 485– 497.
Dai, X. (2009) Thematic and Situational Features of Chinese BBS Texts. Language@Internet, 2009/6. http://www.languageatinternet.org/articles/2009/2011.
Danglli, L. and Abazaj, G. (2014) 'Lexical cohesion, word choice and synonymy in academic writing', in *Mediterranean Journal of Social Sciences*, 5(14), pp. 628-632. http://www.mcser.org/journal/index.php/mjss/article/viewFile/3196/3150.
Davis, J. (2003). 'Expressions of gender: an analysis of pupils' gendered discourse styles in small group classroom discussions', *Discourse & Society,* 14(2), pp. 115-132.
Divjak, D. (2006) 'Ways of intending: Delineating and structuring near synonyms', in Gries, S. T. and Stefanowitsch, A. (eds.) *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis.* Berlin/New York: Mouton de Gruyter, pp. 19–56.
Divjak, D. & Gries, S. T. (2006) 'Ways of trying in Russian: Clustering behavioral profiles', *Corpus Linguistics and Linguistic Theory*, 2 (1), pp. 23–60.
Divjak, D., Arppe, A. and Baayen, H. (in press) 'Does language-as-used fit a self-paced reading paradigm? (The answer may well depend on the statistical model you use)', in Gattnar, A., Anstatt, T. and Clasmeier, C. (eds.) *Slavic Languages in the Black Box*. Tübingen: narr-Verlag.
Dodd, B. (2000) 'Introduction: the relevance of corpora in German studies', in Dodd, B. (ed.) *Working with German Corpora*. Birmingham: University of Birmingham Press, pp. 1–39.
DuBois, P. H. (1939) 'The sex difference on the color-naming test', *American Journal of Psychology*, 52, pp. 380-382.
Eckert, P. (1997) 'Age as a sociolinguistic variable', in Coulmas, F. (ed.) *The Handbook of Sociolinguistics*. Oxford: Blackwell, pp. 151-167.
Edmonds, P. (1999) Semantic representations of near-synonyms. Unpublished thesis. University of Toronto.
Fang, Y. (1989) 'A discussion on the Theme-Rheme structure of Chinese', *Journal of Tshinghua University (Philosophy and Social Sciences)*, 4(2).
Fang, Y., & McDonald, E. (2001) 'On functional structures in Chinese clause' [SIC], *Journal of Foreign Languages*, 1, pp. 42-46.
Firth, J. (1951). 'Modes of meanings' in *Reprinted in Papers in Linguistics 1934-1951*. London: Oxford University Press, pp. 190-215.
Firth, J. (1957). *Papers in linguistics: 1934-1951*. London: Longmans: Green & Co.
Firth, J. (1968) 'A synopsis of linguistic theory, 1930-55', in Palmer F. R. (ed.) *Selected Papers of J R Firth 1952-1959*. Bloomington: Indiana University Press.
Francis, G. (1991) 'Nominal groups and clause structure', *Word* 42 (2), pp. 145-56.
Fries, C. C. (1952) *The structure of English: an introduction to the construction of English sentences.* New York: Harcourt Brace.

Geeraerts, D. (1986) 'On necessary and sufficient conditions', *Journal of Semantics*, 5 (4), pp. 275–291.

Goodman, N. (1952) 'On likeness of meaning', in Linsky, L. (ed.) *Semantics and the philosophy of language: A collection of readings*. Urbana, Illinois: University of Illinois Press, pp. 67-74. Original work published (1949) *Analysis* 13, 1-7.

Gove, P. B. (ed.) (1984) *Webster's New Dictionary of Synonyms.* Springfield, MA: Merriam-Webster.

Gowers, E. (1986). *The Complete Plain Words.* 3rd edn. London: H.M.S.O.

Greenbaum, S. (1969) *Studies in English adverbial usage*. London: Longmans.

Greenbaum, S. (1974) 'Some verb-intensifier collocations in American and British English', *American Speech*, 4, pp. 79–89.

Greenbaum, S. (1996) *The Oxford English Grammar*. Oxford: Oxford University Press.

Gries, S. T. (2001) 'A corpus linguistic analysis of English -ic vs -ical adjectives', *ICAME Journal*, 25, pp. 65–108.

Gries, S. T. & Otani, N. (2010) 'Behavioral profiles: A corpus-based perspective on synonymy and antonymy', *ICAME Journal*, 34, pp. 121–150.

Gries, S. T. (2015) 'Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions', *Language and Linguistics*, 16(1), pp. 93–117.

Halliday, M. A. K. (1959). *The Language of the Chinese 'Secret History of the Mongols'*. Oxford: Blackwell.

Halliday, M. A. K. (1961) 'Categories of the theory of grammar', *Word*, 17/3, pp. 241-292.

Halliday, M. A. K. (1966) 'Lexis as a linguistic level', in Bazell, C., Catford, J., Halliday, M. and Robins, R. (eds) *In Memory of J.R. Firth.* London: Longman, pp.148-162.

Halliday, M. A. K. (1976). 'English System Networks', in Gunther, K. (ed.) *Halliday: system and function in language.* London: Oxford University Press, pp. 101-135.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English.* London: Longman.

Halliday, M. A. K. (1994). *An introduction to functional grammar.* London: Arnold.

Halliday, M. A. K. (1994/2000). *An introduction to functional grammar.* Beijing: Foreign Language Teaching and Research Press.

Halliday, M. A. K. (1996) 'Systemic Functional Grammar', in Brown, K. and Miller, J. (eds.) *Concise Encyclopedia of Syntactic Theories*. Oxford: Pergamon, pp. 321-325.

Hamilton, C., Adolphs, S. & Nerlich, B. (2007) 'The meanings of *risk*: A view from corpus linguistics', *Discourse and Society,* 18 (2), pp. 163–181.

Hanks, P. (1996) 'Contextual dependency and lexical sets', *International Journal of Corpus Linguistics*, 1 (1), pp. 75–98.

Hardie, A. (2012) 'CQPweb - combining power, flexibility and usability in a corpus analysis tool', *International Journal of Corpus Linguistics*, 17(3), pp. 380–409.

Harold, P. and Hornby, A. S. (1933). *The Second Interim Project on English Collocations.* Kaitakusha, Tokyo.
Harris, R. (1973). *Synonymy and linguistic analysis*. Toronto: University of Toronto Press.
Harris, R. A. (1993). *The linguistics wars*. Oxford: Oxford University Press.
Hatim, B. and Mason, I. (1990) *Discourse and the Translator*. Longman.
Herring, S. C. (1993) 'Gender and democracy in computer-mediated communication', *Electronic Journal of Communication*, 3(2). Retrieved June 3, 2003, from http://www.cios.org/getfile/HERRING_V3N293.
Hirsch (1975) 'Stylistics and Synonymity', *Critical Inquiry*, 1(3), pp. 559-579. University of Chicago Press. http://www.jstor.org/stable/1342831
Hoey, M. (1991) *Patterns of lexis in text*. Oxford: Oxford University Press.
Hoey, M. (2004) 'The textual priming of lexis', in Aston, G., Bernardini, S. & Stewart, D. (eds.) *Corpora and Language Learners*. Amsterdam: John Benjamins Publishing, pp. 21-41.
Hoey, M. (2005). *Lexical Priming: a new theory of words and language*. London: Routledge.
Hoey, M. (2007) 'Lexical priming and literary creativity', in Hoey, M., Mahlberg, M. Stubbs, M. and Teubert, W. (eds) *Text, Discourse and Corpora Theory and Analysis*. London: Continuum, pp. 7-30.
Hoey, M. and Shao, J. (2015) 'Lexical priming: The odd case of a psycholinguistic theory that generates corpus-linguistic hypotheses for both English and Chinese', in Zou, B., Hoey, M. and Smith, S. (eds) *Corpus Linguistics in Chinese Contexts*. New York: Palgrave Macmillan, pp. 15-34.
Holmes, J. (1986) 'Compliments and compliment responses in New Zealand English', *Anthropological Linguistics*, 28(4), pp. 485-508.
Hounston, A. (1989) 'The English gerund: syntactic change and discourse', in Fasold, R.W. and Schiffrin, D. (eds.) *Language change and variation*. John Benjamins Publishing, pp. 173-196.
Hu, Z-L. (1994) *Discourse cohesion and coherence.* Shanghai: Shanghai Foreign Language Education Press.
Hu, X-J. (1997) *The thematic structure of modern Chinese*. Unpublished Master's thesis. Sichuan Foreign Studies University.
Huang, Ch.-R. and Hong, J.-F. (2005) 'Deriving conceptual structures from sense - A study of near synonymous sensation verbs', *Journal of Chinese Language and Computing* 15(3), pp. 125-135.
Hunston, S. (2001) 'An empirical grammar of the English verb system', *System* 29(2), pp. 313–316.
Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.
Hunston, S. (2007) 'Semantic prosody revisited', *International Journal of Corpus Linguistics*, 12(2), pp. 249–68.
Hüllen, W. (2004). *A history of Roget's thesaurus: origins, development, and design*. Oxford: Oxford University Press.
Hüllen, W. (2009) 'A cognitive view of synonymy', in *Networks and Knowledge in Roget's*

Thesaurus. Online: University Press Scholarship Online.
Jesperson, O. (1949) *A Modern English Grammar on Historical Principles*. London: Allen & Unwin.
Jones, S. (2002). *Antonymy: a corpus based perspective*. London: Routledge.
Justeson, J. S. and Katz, S. M. (1995) 'Principled disambiguation: Discriminating adjective senses with modified nouns', *Computational Linguistics*, 21 (1), pp. 1–27.
Kempson, R. (1977). *Semantic Theory*. Cambridge: Cambridge University Press.
Kennedy, G. (1998). *An Introduction to Corpus Linguistics.* Longman, London and New York.
Kilgarriff, A. (2003). *The sketch engine*. http://www.sketchengine.co.uk
Kilgarriff, A. and Kosem, I. (2013) 'Corpus evidence and electronic lexicography', in Granger, S. and Paquot, M. (eds) *Electronic Lexicography*. Published to Oxford Scholarship Online. DOI:10.1093/acprof:oso/9780199654864.001.0001
Krenn, B. and Stefan, E. (2001) 'Can we do better than frequency? A case study on extracting PP-verb collocations', in *Proceedings of the ACL/EACL Workshop on the Computational Extraction, Analysis and Exploitation of Collocations*. Toulouse, France, pp. 39–46.
Krishnamurthy, R. (2000) 'Collocation: from *silly* ass to lexical sets', in Heffer C., Sauntson H. and Fox G. (eds.) *Words in Context: A tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham.
Lakoff, R. (1975). *Language and Woman's Place.* New York: Harper & Row.
Lakoff, G. and Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.
Lakoff, G. (1993) 'The contemporary theory of metaphor', in *Metaphor and Thought*. UC Berkeley: Retrieved from: http://escholarship.org/uc/item/54g7j6zh
Leech, G. (1981) *Semantics: The Study of Meaning*. Harmondsworth: Penguin.
Leech, G. (1991) 'The state of the art in corpus linguistics', in Aijmer, K. and Altenberg, B. (eds.) *English corpus linguistics*. London: Longman, pp. 8-29.
Leech, G. (1997) 'Teaching and Language Corpora: a convergence', in Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.) *Teaching and Language Corpora*. London and New York: Longman, pp, 1-23.
Lee, Ch-Y. and Liu, J.-S. (2009) 'Effects of collocation information on learning lexical semantics for near synonym distinction', *Computational Linguistics and Chinese Language Processing*, 14(2), pp. 205-220.
Li, C. N. and Thompson, S. A. (1989). *Mandarin Chinese: A Functional Reference Grammar.* University of California Press.
Li, S-H. (2007). *A systemic functional grammar of Chinese*. London: Continuum.
Liu, M.-Ch., Huang, C.-R., Lee, C. and Lee, Ch.-Y. (2000) 'When endpoint meets endpoint: A corpus-based lexical semantic study of Mandarin verbs of *throwing', Computational Linguistics and Chinese Language Processing*, 5(1), pp. 81-96.

Liu, D. (2010) 'Is it a chief, main, major, primary, or principal concern: A corpus-based behavioral profile study of the near-synonyms', *International Journal of Corpus Linguistics*, 15(1), pp. 56–87.
Liu, D and Espino. M. (2012) 'Actually, genuinely, really, and truly: A corpus-based behavioral profile study of near-synonymous adverbs', *International Journal of Corpus Linguistics* 17(2), pp. 198-228.
Louw, B. (1993) 'Irony in the text or insincerity in the writer? -- The diagnostic potential of semantic prosodies', in Baker, M., Francis, G. and Tognini-Bonelli, T. (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, pp. 157-176.
Lu, J. (2010) 'A corpus-based study on collocational behavior and semantic prosody of near synonyms in Chinese learner English', *Modern Foreign Languages*, 2010(03).
Lyons, J. (1968). *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press.
Lyons, J. (1977). *Semantics*. 2 vols. Cambridge: Cambridge University Press.
Lyons, J. (1981). *Language, meaning, and context.* Cambridge: Cambridge University Press.
Lyons, J. (1995). *Linguistic semantics: an introduction*. Cambridge: Cambridge University Press
Mackay, S. (1980) 'Teaching the syntactic, semantic, and pragmatic dimensions of verbs', *TESOL Quarterly*, 14, pp. 17-26.
Manning, C. and Schütze, H. (2001) *Foundations of Statistical Natural Language Processing.* Cambridge: MIT Press.
Martin, M. (1984) 'Advanced vocabulary teaching: The problem of synonyms'. *The Modern Language Journal*, 68(2), pp. 130-137.
McEnery, T. and Wilson, A. (2001) *Corpus Linguistics: An introduction*. Edinburgh: Edinburgh University Press.
McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies*. London/New York: Routledge.
McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
Meyer, D. E. and Schvanefeldt, R.W. (1971) 'Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations', *Journal of Experimental Psychology*, 90(2), pp. 227–34.
Miller, G. A. & W. G. Charles. (1991) 'Contextual correlates of semantic similarity', *Language and Cognitive Processes*, 6, pp. 1-28.
Nissen, H. B. and Henriksen, B. (2006) 'Word class influence on word association test results', *International Journal of Applied Linguistics*, 16(3), pp. 389–408.
Nuyts, J. (2012). 'Notions of (inter)subjectivity', *English text construction*, 5(1), p.53-76.
Oxford Dictionary Online. http://www.oxforddictionaries.com/
Pace-Sigge, M. (2013). *Lexical priming in spoken English usage*. Basingstoke: Palgrave Macmillan.
Palmer, F. R. (1976) *Semantics: A new outline*. Cambridge: Cambridge University Press.

Palmer, F. R. (1981). *Semantics.* 2nd edn. Cambridge: Cambridge University Press.
Pan, F. (2010) 'Lexical acquisition viewed from a contrastive analysis of collocational behavior of near synonyms', *Chinese Journal of Applied Linguistics (Bimonthly)*, 33, pp. 52-64.
Partington, A. (1998). *Patterns and meanings: using corpora for English language research and teaching.* Amsterdam; Philadelphia: J. Benjamins.
Partington, A. (2003). *The linguistics of political argument: the spin-doctor and the wolf-pack at the White House*. London; New York: Routledge.
Partington, A. (2004) 'Utterly content in each other's company: Semantic prosody and semantic preference', *International Journal of Corpus Linguistics,* 9(1), pp. 131-156.
Pearce, D. (2002) 'A comparative evaluation of collocation extraction techniques', *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*. Las Palmas, Canary Islands, pp. 1530-1536
Pearson, J. (1998). *Terms in context*. Amsterdam; Philadelphia: J. Benjamins.
Plato. (1953) 'Laches', in Jowett, B. (tr.) *The Dialogues of Plato*. 4th edn. Oxford: Clarendon Press, pp. 67-102.
Postman, L. and Keppel, G. (1970). *Norms of Word Associations*. New York: Academic Press.
Pustejovsky, J. (1996). *The Generative Lexicon*. 2nd print. Cambridge MA: MIT Press.
Quine, W.V.O. (1951) 'Two dogmas of empiricism', *The Philosophical Review*, 60, pp. 20-43.
Quine, W.V.O. (1953/1980). *From a Logical Point of View*. 2nd edn. Cambridge, MA: Harvard University Press.
Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Langugae*. London: Longman.
Ramisch, C., Schreiner, P., Idiart, M. and Villavicencio, A. (2008) 'An Evaluation of Methods for the extraction of Multiword Expressions', *Proceedings of the LREC-2008 Workshop on Multiword Expressions: Towards a Shared Task for Multiword Expressions*. Marrakech, Morocco, pp. 50-53.
Rich, E. (1977) 'Sex-related differences in colour vocabulary', *Language and speech*, 20(4), pp. 404 - 409, Carnegie-Mellon University.
Riemer, R. (2010). *Introducing Semantics.* Cambridge: Cambridge University Press.
'Priming' (2016) *Wikipedia.* Available at https://en.wikipedia.org/wiki/Priming_(psychology) (accessed on 30th April, 2016).
Rychlý, P. (2008) 'A Lexicographer-Friendly Association Score', in Sojka, P. and Horák, A. (eds.) *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008*. Masaryk University, Brno, pp. 6–9.
Schmied, J. (1993). 'Qualitative and quantitative research approaches to English relative constructions', in *Souter, C and Atwell, E.* (eds) *Corpus-based Computational Linguistics*, pp. 85-96.
Sinclair, J. (1987) 'Collocation: A progress report', in Steele, R. and Tomas, T. (eds.) *Language*

Topics: Essays in Honor of Michael Halliday, II. Amsterdam/Philadelphia: John Benjamins, pp. 319–331.

Sinclair, J. (1991a). *Corpus, Collocation, Concordance.* Oxford University Press, Oxford.

Sinclair, J. (1991b). 'Shared knowledge', in *Georgetown University Round Table on Languages and Linguistics.* Washington D.C.: Georgetown University Press, pp. 489-500.

Sinclair, J. (1992) 'Priorities in discourse analysis', in Coulthard, M. (ed.) *Advances in Spoken Discourse Analysis*. London: Routledge, pp. 79-88.

Sinclair, J. (1996) 'The search for units of meaning', *TEXTUS IX*, pp. 175-106.

Sinclair, J. (1999) 'The lexical item', in Weigand, E. (ed.) *Contrastive Lexical Semantics*. Amsterdam/Philadelphia: John Benjamins, pp. 1-24.

Sinclair, J. (2004). (Edited with Carter, R.) *Trust the Text: Language, Corpus, and Discourse*. London: Routledge.

Sinclair, J. et al. (2005) *Collins COBUILD English grammar*. Glasgow: HarperCollins.

Sluiter, I. (1993) 'The Greek tradition', in Bekkum, W. et al (eds.) *The emergence of semantics in four linguistic traditions.* Amsterdam: Benjamins, pp.147-224.

Starcke, B. (2008) 'I don't know -- differences in patterns of semantic prosody in phrases of different lengths', in Gerbig, A and Mason, A. (eds) *Collocation and Language, People, Numbers*. Amsterdam, NL: Rodopi, pp. 199-216.

Storjohann, P. (2010) 'Synonyms in corpus texts: conceptualisation and construction', in Storjohann, P. (ed) *Lexical-Semantic Relations: Theoretical and Practical Perspectives.* Amsterdam, NLD: John Benjamins Publishing Company, pp. 69-95.

Stubbs, M. (1993) 'British traditions in text analysis from Firth to Sinclair', in Baker, M., Francis, G. and Tognini-Bonelli, E. (eds) *Text and Technology: In Honour of John Sinclair*. Amsterdam, NLD: John Benjamins Publishing Company, pp. 1-34.

Stubbs, M. (1995) 'Collocations and semantic profiles: On the cause of the trouble with quantitative methods', *Function of Language*, 2(1), pp. 1–33.

Stubbs, M. (1996). *Text and corpus analysis: computer-assisted studies of language and culture.* Oxford: Blackwell.

Stubbs, M. (2001). *Words and phrases.* Oxford: Blackwell.

Sun, K.-T., Huang, Y.-M., and Liu, M.-C. (2011) 'A WordNet-based near-synonyms and similar-looking word learning system', *Educational Technology & Society,* 14 (1), pp. 121–134.

Sweetser, E. (1990) *From Etymology to Pragmatics*. Cambridge: Cambridge University Press.

Sylvia, C. and Weiner, E. 91998) *Oxford dictionary of English grammar*. New York: Oxford University Press

Tantucci, Vittorio. (2017). 'From immediate to extended intersubjectification: a gradient approach to intersubjective awareness and semasiological change', *language and cognition*, 9(1), pp. 88-120.

Doi:10.1017/langcog.2015.26.

Taylor, J. (2003). *Linguistic categorization.* 3rd edn. Oxford: Oxford University Press.

Taylor, R. (2003) 'Near synonyms as co-extensive categories: ''high'' and "tall" revisited', *Language Sciences*, 25, pp. 263-284.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: John Benjamins.

Violi, P. (2001). *Meaning and experience*. Indiana: Indiana University Press.

Wang, H. H. & Wang, T. S. (2005) 'CAUSE 语义韵的对比研究 [A contrastive study on the semantic prosody of CAUSE]', *Modern Foreign Languages*, 3, pp. 297-307.

Wang, T. and Hirst, G. (2010) 'Near-synonym Lexical Choice in Latent Semantic Space', *Proceedings of the 23rd International Conference on Computational Linguistics.* Beijing.

Wei, N. X. (2006) '基于语料库学生英语中的语义韵对比研究 [A corpus-based contrastive study of semantic prosodies in learner English]', Foreign Language Research, 5, pp. 50-54.

Wen, L. X. (2007) '基于语料库数据的近义词语义韵调查 [Corpus data-based study of semantic prosody of synonyms]', *Journal of Shenyang University*, 5, pp. 56-60.

Wierzbicka, A. (1987) *English Speech Acts Verbs: A Semantic Dictionary.* New York: Academic Press.

'Metonymy' (2016) *Wikipedia*. Available at https://en.wikipedia.org/wiki/Metonymy (accessed on 13th Jan, 2016).

Wolfson, N. (1983) 'An empirically based analysis of complimenting in American English', in Wolfson, N. and Judd, E. (eds) *Sociolinguistics and Language Acquisition*. Rowley, Mass: Newbury House, pp. 82-95.

Wu, P.-Y., Chen, P.-H. and Gong, Sh.-P. (2011) 'Collocation, Semantic Prosody and Near Synonymy: Verbs of ''Help'' and ''Aid'' in Mandarin Chinese', *12th Chinese Lexical Semantics Workshop.*

Xiao, R. and McEnery, T. (2006) 'Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective', *Applied Linguistics,* 27(1), pp. 103-129.

Zoltan, K. (2010) *Metaphor: A Practical Introduction* (2nd edn). Cary, NC, USA: Oxford University Press.

蔡美智.(2010)「同様、相同」不「一様」：表相似近義詞指稱功能辨析。華語文教學研究，第七卷第一期: 57-59。